ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

CONTENTS

VOL. 6, NO. 3 — JULY, 1938

ECONOMETRICA

Published quarterly by the Econometric Society
Colorado Springs, Colorado, U. S. A.

BUSINESS OFFICE OF ECONOMETRICA, Mining Exchange Building, Colorado Springs, U. S. A. In the absence of specifications to the contrary, any communication regarding ECONOMETRICA or the Econometric Society, may be addressed to this office.

Subscription Price $7.00 per annum. Single copies $1.75.

Printed in the U. S. A. by George Banta Publishing Company
Menasha, Wisconsin

ECONOMETRICA

VOLUME 6 JULY, 1938 NUMBER 3

TRAGEDIES IN THE LIFE OF COURNOT*

By A. J. NICHOL

ANDREW CARNEGIE once said, "It does not pay to pioneer." Certainly, if there is any truth in this statement, it finds ample illustration in the life and work of Antoine-Augustin Cournot. The pioneering which he did brought him no financial reward, little academic recognition, and scarcely any intellectual comradeship. Alfred Marshall in his old age received a substantial income from royalties on his books, particularly his *Principles of Economics;* but the book of which we celebrate the hundredth anniversary of original publication brought its author no such return. In a letter to Walras in 1873 Cournot confessed that he was very unpopular with his publishers because none of his books found any considerable number of readers until years after their first appearance.[1]

The tragedies in the life of Cournot were not tragedies of stark physical misery such as have so often attended the lives of geniuses in literature, art, and philosophy. Cournot did not starve in a garret. He was not carried off by death in the springtime of youth as were Keats, Shelley, Byron, Abel, and Galois. He was not picked up out of a gutter to die as was Edgar Allan Poe. Economists generally do not do that sort of thing. The tragedies to which economists are usually subject are intellectual rather than physical. They are tragedies of the soul. So it was in the main with Cournot. Yet, as we shall see, a physical infirmity lay at the root of many of his intellectual disappointments.

Among us in America and England Cournot is known as an economist, but in France he is chiefly remembered as a philosopher. The article on Cournot in the *Encyclopedia of the Social Sciences,* written by one of his own countrymen, gives much more attention to his philosophical writings than to his work in economics.[2] So it was also in the

* Paper read at the Cournot Memorial session of the Econometric Society, Wednesday, December 29, 1937, at its Atlantic City meeting. Antoine-Augustin Cournot was born in 1801 and died in 1877. In 1838 he published his pioneer work. The above session was held in recognition of the centennial of its appearance.

[1] Letter of Cournot to Walras, dated Sept. 3, 1873, reprinted in ECONOMETRICA, Vol. 3, January, 1935, pp. 119–120.

[2] C. Bouglé, "Cournot, Antoine-Augustin," in E. R. A. Seligman, Editor-in-Chief, *Encylopedia of the Social Sciences*, New York, 1931, Vol. 4, p. 511.

obituaries[3] published in Paris after his death in 1877. When a movement was started in 1905 for reconsideration of his work, the results were gathered together in no economic publication but in a philosophical journal.[4] Among those with whom he mingled in day-to-day routine, however, Cournot was known as an administrative official of the French public schools. Not the least of the legacies left to France by Napoleon was its highly centralized educational system. Of it long afterward one Minister of Education boasted he could tell exactly what recitations were being carried on at any hour in any classroom. Cournot's position in this system was one of considerable importance. His duties corresponded in many ways to those of an American school superintendent, but were much broader in scope, extending not only to elementary and secondary schools but also to provincial branches of the University of France. The men who occupied such positions were, as they are in France today, a select group of great ability, burdened with responsibility, but not particularly well paid. Cournot's own salary probably never exceeded 15,000 francs per annum. Yet in one important position he was the immediate successor of Ampère, the physicist whose name has been made immortal as the designation of a unit of electrical measurement.

Cournot was born in 1801. In his early years he sought the companionship of adults rather than children his own age. Uncles, aunts, and grandparents, and elderly servants were his chief associates. He read voraciously in spite of a premonition that he might lose his eyesight—a premonition which was almost literally fulfilled. After local schooling he became interested in continuing his education in Paris, and at the age of twenty enrolled in the then existing *École normale*. The school was abruptly disbanded at the end of his first year. Half the students were immediately appointed to positions in the school system; the others were allowed a very small temporary pension. Cournot found himself in the latter group, being considered lacking in piety. The suspicion was most unjust for Cournot was always a loyal, though unostentatious, son of the Church. Nevertheless, this turn in affairs ultimately proved to be a very favorable one. If he had been sent out into the rural districts at this time to explain "the square on the hypotenuse," he probably would have spent the rest of his life in utter obscurity. As it was, he remained in Paris, and, when his pension expired, found employment as secretary to one of the generals of Napoleon, the Marshal Gouvion Saint-Cyr. Thus he was enabled to continue his studies at the University, complete the requirements for a doctor's

[3] *Revue des deux mondes*, Vol. 22, 1877, pp. 102–104; *Journal des économistes*, Vol. 46, 1877, pp. 304–305, reprinted from *Le Temps*.

[4] *Revue de métaphysique et de morale*, 13[e] année, No. 3, May, 1905, *numero specialement consacré a Cournot.*

degree, and come in contact with some of the most inspiring intellectual leaders of the time. For ten years, however, i.e., 1823–1833, this genius of economics and philosophy was principally occupied in editing military memoirs. His first printed work of any length was a hundred-page biography of his soldier-employer.[5]

Some mathematical articles, written by Cournot as a University student, attracted the attention of Poisson, the physicist, who occupied a position of power in educational circles. With Poisson as his mentor Cournot was finally placed in the educational system of France, and made rapid advancement—far more in all probability than he could ever have attained had he entered the system ten or twelve years earlier in accordance with his original expectations. He was first appointed professor of mathematics at Lyons in 1834. His principal course was in differential calculus. Since the course and the professor were both entirely new, the lecture hall was thronged the first few days; but, perceiving the difficulty of the subject, most of the students departed, leaving Cournot to carry on with a very small group. This year at Lyons was the only year in which Cournot was directly engaged in teaching. In other directions it was very productive for in 1834–35 the youthful professor of mathematics completed the major portion of his first and greatest work in economics and the main outlines of his *Exposition de la théorie des chances et des probabilités.*[6] The very next year at the age of thirty-four he was appointed Superintendent of the school district (*Recteur de l'academie*) of Grenoble. Familiarity with the region helped him become a successful administrator. Within a few months his responsibilities were increased by assumption of the duties of Inspector General of Education in addition to those of *Recteur.* As Inspector General he took the place temporarily of Ampère, who had suddenly died.

The year we now celebrate was a very eventful one in the life of Cournot. In the midst of the onerous routine of 1838 he was married, published his *Recherches dans les principes mathématiques de la théorie des richesses*, and received another promotion, being assigned permanent duties as one of several travelling Inspectors General of Education with headquarters in Paris. With great regret he left Grenoble which had seemed to offer him so much peace, contentment, and opportunity for accomplishment.

His friend, Poisson, died in 1840. Thereafter Cournot made no further spectacular progress in the service of the University. Trouble with his eyes forced him in 1844 to take a year's leave of absence, and this he spent in Italy. He had experienced difficulties with vision while he was

[5] In Maréchal Gouvion Saint-Cyr, *Mémoires pour servir a l'histoire militaire*, 4 vols., Paris, 1831.

[6] Published in Paris, 1843.

a student in the University and secretary to Marshal Gouvion Saint-Cyr. This condition had gradually grown worse through the years. After his return from Italy he does not seem to have been seriously handicapped in the performance of his regular duties as administrator and supervisor, but there was a very profound change in the character of his published work. To Walras in 1873 he wrote,

"Thirty years ago I had to renounce all mathematics."[7]

It became impossible for him to engage in any long-continued close work with his eyes. He did not give up his companionship with books. Secretaries read aloud to him regularly for hours at a time. He never found any one, however, who could read mathematics to him. Symbols, the meaning and possibilities of which he had once grasped so masterfully through the eye, he was never able to comprehend by ear. When Professor Irving Fisher edited the English translation of Cournot's *Researches in the Mathematical Principles of the Theory of Wealth,* he found the original French version replete with mistakes. Some of them were trivial and easily corrected; others were fundamental errors in mathematical reasoning. Thus Cournot's great work bore the appearance of gross carelessness in proofreading and original composition.[8] Such perhaps there may indeed have been. Another very pertinent explanation of Cournot's errors of 1838, however, lies in the fact that even then he was approaching a state of partial blindness.

To the academy (school district) of Dijon Cournot came as Rector (Superintendent) in 1854, having declined another more lucrative appointment because of political complications. He remained at Dijon until his retirement in 1862. Then, vigorous otherwise but eyesight worse than ever, almost Milton-like, he made his home again in Paris.

A few years before retirement Cournot completed a very illuminating volume of personal memoirs. They remained in manuscript form, almost entirely unknown, for more than fifty years before they were published.[9] These memoirs, and the voluminous philosophical works which Cournot had printed both before and after his retirement,[10]

[7] See note 1.

[8] Irving Fisher, "Cournot and Mathematical Economics," *Quarterly Journal of Economics*, Vol. 12, 1898, p. 129.

[9] A. Cournot, *Souvenirs 1760–1860*, (E. P. Bottinelli, Ed.), Paris, 1913. This is the original source of otherwise undocumented material in the text above. Professor William Jaffé of Northwestern University is, as far as I am able to discover, the owner of the only copy in the United States. A summary of the *Souvenirs* appears in French in H. L. Moore, "Antoine-Augustin Cournot," in *Revue de métaphysique et de morale*, May, 1905, pp. 521–543.

[10] *Essai sur les fondaments de nos connaissances et sur les caracteres de la critique philosophique*, 1851; *Traité de l'enchaînment des idées fondamentales*, 1861; *Considérations sur la marche des idées et des evénements dans les temps modernes*, 1872; *Matérialisme, vitalisme, rationalisme*, 1875.

were written in a very curious way. They were first jotted down piecemeal on scraps of paper, then collected and copied by an amanuensis, read back to him, and corrected.

Outside of a few modest advertisements by his publisher the first printed reference to Cournot's work in economics came in a twenty-page review[11] in the *Journal des Économistes*, August, 1864. In the first few months of his retirement Cournot had written a second book in economics.[12] Its appearance turned the attention of the reviewer back to Cournot's first book, written twenty-five years previously; and to the earlier work the review was principally devoted. In conclusion Cournot was upbraided for lack of progress and lack of familiarity with other work in economics. Obstinate in his independence, Cournot in the preface to his third book in economics (published after his death) replied, saying in effect:

"I am the only French economist who has not been cited by the others. I'll be d——d if I cite them."[13]

Fifteen years Cournot lived in retirement on the pension of an old employee of the University. These years were tinged with bitterness and disappointment. The news came that another young French professor, Leon Walras, was applying to economics the same technique which Cournot himself had used. Alas, it was a technique which had long before been renounced by its originator. As the shadows darkened, a possibility appeared of his election to the French Academy as an economist. To further his chances his friends urged him to write another book, and this he did; but death suddenly took him away without receipt of further honor. He did not live to see his last book in print.[14]

To us who follow in the footsteps of Cournot one tragedy of his life is that in his most productive years he was so much occupied with administrative work. This has always been a supreme obstacle in the way of individual scholarship. If he had been able to settle down to teaching and writing while still in his prime, how much more he might have accomplished no one knows. The supreme tragedy of Cournot's life without doubt was the tragedy of blindness. If he had found a good occulist, if he had been a well man, he might have advanced the progress of mathematical economics a generation. It would not then have been left to a Marshall, a Walras, and Irving Fisher to make his ideas known to the world. He might have done it himself.

University of California
Berkeley, California

[11] *Journal des économistes*, Vol. 43, 1864, pp. 231–251. The reviewer was R. de Fontenay.

[12] *Principes de la théorie des richesses*, Paris, 1863.

[13] *Revue sommaire des doctrines économiques*, Paris, 1877, p. iii.

[14] *Revue de métaphysique et de morale*, May, 1905, p. 345.

COURNOT FORTY YEARS AGO*

By Irving Fisher

It may be of interest to members of the Econometric Society on this one-hundredth anniversary of the publication of Cournot's *Recherches sur les principes mathématiques de la théorie des richesses* to recall what was said of this epoch-making work forty-one years ago when it was translated into English.

The translation[1] was made by the late Nathaniel T. Bacon, my wife's brother-in-law, with a "Bibliography of Mathematical Economics" by myself and later an article on the book, in the *Quarterly Journal of Economics*, which was also by me. This appeared in January, 1898, and was entitled "Cournot and Mathematical Economics"; it was followed by an Appendix of mathematical notes, the object of which was to help the reader through Cournot's rather condensed mathematical reasoning.

As was stated in the preface to the translation,

> The bibliography divides itself very naturally into four periods, beginning respectively with the treatises of Ceva, Cournot, Jevons, and Marshall. Ceva enjoys the distinction of being the first known writer to apply mathematical method to economic problems; Cournot was apparently the first to apply it with any great degree of success; Jevons (and almost simultaneously Walras) attracted the serious attention of economists to this method; and Marshall brought it (or at any rate its simpler diagrams) into general use. The four periods are of constantly decreasing length, being respectively, 127, 33, 19, and 8 years, but the number of titles grows greater in each succeeding period.

It may be of interest also to quote the last sentence of the preface to the translation:

> Grateful acknowledgments are due to the many other persons who have supplied bibliographical data, and in particular to Professors Pantaleoni, Walras, Pareto, and Edgeworth.

It is significant that these four voices from the past are listened to today with greater respect than when they were thus cited in 1897 thanks to the progress since then of the mathematical method for which they, following Cournot, stood.

Mathematical economics, now so generally recognized, was still looked at askance in 1897 when the English edition of Cournot appeared. This had been true for many years. Ingram, in his *History of Political Economy*, 1888, had said of mathematical economics (p. 182),

* Paper read at the Cournot Memorial session of the Econometric Society, Wednesday, December 29, 1937, at its Atlantic City meeting.

[1] *Mathematical Principles of the Theory of Wealth*, N. Y. (Macmillan) in the "Economic Classics Series," edited by Professor W. J. Ashley.

There is, therefore, no future for this kind of study; and it is only waste of intellectual power to pursue it.

Nevertheless it was even then showing signs of a healthy growth. As was stated in the article on Cournot in the *Quarterly Journal of Economics:*

For good or for ill the mathematical method has finally taken root, and is flourishing with a vigor of which both its friends and enemies little dreamed. Sixty years ago the mathematical treatise of Cournot was passed over in silence, if not contempt. To-day the equally mathematical work of Pareto is received with almost universal praise. In Cournot's time "mathematical economists" could be counted on one's fingers, or even thumbs. To-day [1898] they muster some thirty active enthusiasts and a much larger number of followers and sympathizers. In 1838 there seems to have been no institution of learning besides the Academy at Grenoble, of which Cournot was rector, where "mathematical economics" was employed or approved. In 1898 there are at least a dozen such institutions, and in England alone half that number, Oxford and Cambridge among them. It is in France, the prophet's own country, where he is still without honor. When Cournot wrote, no journal existed in which such investigations as his could find a welcome. To-day the *Economic Journal*, the *Journal of the Royal Statistical Society*, the *Giornale degli Economisti*, and the *Nationaloekonomisk Tidsskrift* receive such material with more or less regularity; while, within the last eight years alone, twenty other journals have occasionally published economic articles containing mathematics. Opponents of the new method no longer venture to ignore or ridicule it, but, in academic circles at least, seek to acquaint themselves with its history and present aims as matters of necessary and professional information. In recognition of such widespread interest the latest *Dictionary of Political Economy* devotes some forty articles to the history, writings, methods, and terminology of the "mathematical school."

It may fairly be claimed that Cournot was the principal founder of this school.[2] For this reason, if for no other, his book is an "economic classic," and as such deserves careful study. But its interest is not simply historical. The bulk of its reasoning and conclusions has never yet been superseded. Those who now read it for the first time will find it as new and fresh as any modern investigation. As the original work has long been out of print and scarce in the antiquarian market, the present edition serves the double purpose of translation and second edition. Moreover, thanks to the painstaking work of the translator, it far surpasses the original in typographical accuracy, a prime requisite in a mathematical work.

During the last forty years both the appreciation of Cournot and the appreciation of mathematical economics in general have grown amazingly and especially in the last five years since the Econometric Society was founded.

To quote further:

The book itself, naturally falls under three heads: introductory chapters, treating of value, "absolute and relative," and of the foreign exchanges, are quite apart

[2] Cf. Walras, *Theorie Mathématique de la Richesse Sociale*, 1883, p. 9.

from the rest of the book. Chapters IV–X inclusive discuss the determination of prices under different conditions as to monopoly and competition, taxes and bounties. This portion of the work is the most distinctive and the most widely celebrated. The remaining two chapters give an ambitious but erroneous theory of "Social Income."

Chapter I is devoted to defining wealth, which term Cournot uses in the sense of value in exchange. He carefully distinguishes this idea from *utility*, with which he conceives the economist has no direct concern. Here of course, he differs materially from modern mathematical economists, beginning with Jevons and Walras.

Cournot was, I believe, the first to make a mathematical study of imperfect competition such as has recently been so highly developed by Professor Zeuthen, Mrs. Robinson, Professor Chamberlin, and others. His treatment was vulnerable but suggestive and brilliant in its way.

It is interesting to find that Cournot made some serious errors, as stated in the article:

In introducing the subject of import duties or bounties, "without pretending, which would be absurd, to contradict the opinion which has been very generally formed, of the advantages for the community procured by improvements in the means of communication or by the extension of markets,"[3] Cournot suggests that the extreme position of free traders is untenable. In following out this contention, Cournot commits a mathematical blunder which invalidates his main thesis; namely, that a tariff on imports may, under certain peculiar circumstances, lower prices of the goods imported. Formulae (6) on page 122 are erroneous[4] for reasons explained in the appended notes (No. 50). The correct formulae may be transcribed from those given by putting zero for ϵ. With this change it will be seen that Cournot's arguments on pages 123 and 124 are quite destroyed.

This singular error supplies one of many examples of a serious fault in our talented author—gross carelessness.[5] In spite of extraordinary acuteness and precision of mind, Cournot was neglectful of his duties as verifier and proof reader. The translator, Mr. Bacon, has convicted him of some thirty-five inaccuracies.

It is interesting to observe that Cournot, like Mill, was not only an economist but also a philosopher, with a fundamental knowledge of physical science. It is not strange, therefore, that he should have made reference in his *Principles* to "absolute" as distinct from relative valuation and to have cited Sir Isaac Newton in support of an absolute space with the implication that absolute value may similarly be valid.

[3] Cournot, *op. cit.*, p. 121.

[4] Cf. Edgeworth, in Palgrave's *Dictionary*, article "Cournot" and Berry and Sanger, quoted by Edgeworth, *Economic Journal*, 1894, p. 627.

[5] When my paper was originally published, I was unaware of the fact that Cournot was suffering from bad eyesight or near blindness and that this probably accounts for such a multitude of errors.

All this is out of tune with modern relativity but not out of tune with the best thought of Cournot's day.

He says:

Finally, there are some circumstances which may make it certain that relative or apparent movements come from the displacement of one body and not of another.[6] Thus the appearance of an animal will show by unmistakable signs whether it is stopping or starting. Thus, to return to the preceding example, experiments with the pendulum, taken in connection with the known laws of mechanics, will prove the diurnal motion of the earth; the phenomenon of the aberration of light will prove its annual motion; and the hypothesis of Copernicus will take its place among established truths.

Let us now examine how some considerations perfectly analogous to those which we have just considered, spring from the idea of exchangeable values.

Just as we can only assign situation to a point by reference to other points, so we can only assign value to a commodity by reference to other commodities. In this sense there are only relative values. But when these relative values change, we perceive plainly that the reason of the variation may lie in the change of one term of the relation or of the other or of both at once; just as when the distance varies between two points, the reason for the change may lie in the displacement of one or the other or both of the two points. . . .

We can therefore readily distinguish the relative changes of value manifested by the changes of relative values from the absolute changes of value of one or another of the commodities between which commerce has established relations.

Just as it is possible to make an indefinite number of hypotheses as to the absolute motion which causes the observed relative motion in a system of points, so it is also possible to multiply indefinitely hypotheses as to the absolute variations which cause the relative variations observed in the values of a system of commodities.

However, if all but one of the commodities preserved the same relative values, we should consider by far the most probable hypothesis, the one which would assign the absolute change to this single article; unless there should be manifest such a connection between all the others, that one cannot vary without involving proportional variations in the values of those which depend on it.

For instance, an observer who should see by inspection of a table of statistics of values from century to century, that the value of money fell about four-fifths towards the end of the sixteenth century, while other commodities preserved practically the same relative values, would consider it very probable that an absolute change had taken place in the value of money, even if he were ignorant of the discovery of mines in America. On the other hand, if he should see the price of wheat double from one year to the next without any remarkable variation in the price of most other articles or in their relative values, he would attribute it to an absolute change in the value of wheat, even if he did not know that a bad grain harvest had preceded the high price.

Without reference to this extreme case, where the disturbance of the system of relative values is explained by the movement of a single article, it is evident that among all the possible hypotheses on absolute variations some explain the relative variations more simply and more probably than others.[7]

[6] See Newton, *Principia*, Book I, at the end of the preliminary definitions.

[7] Cournot, *op. cit.*, pp. 20–22.

Cournot goes on for some ten pages in all in his discussion of "absolute" value. After all, despite Cournot's old-fashioned use of the term "absolute," his observations, if clothed in modern terms of relativity, relating values to the general economic framework, fit in with modern thought just as does Isaac Newton's old-fashioned "absolute space" when translated into terms relative to the general astronomic framework of the "fixed" stars.

Whatever his errors, Cournot did epoch-making work and will always be recognized for it.

To such pioneers homage is due alike for their success in path-breaking and for the courageous trial and error by which alone that success could be attained.

Yale University
New Haven, Conn.

THE METHOD OF SUPPLEMENTARY CONFLUENT RELATIONS, ILLUSTRATED BY A STUDY OF STOCK PRICES[1]

By TRYGVE HAAVELMO

I. DIFFERENT TYPES OF DYNAMIC RELATIONS

1. The Difference Between Structural and Confluent Relations

Let

$$(1) \qquad x_i(t) = \sum_{j \neq i} a_{ij} x_j(t) + \sum_j b_{ij} \dot{x}_j(t) + \sum_j c_{ij} \ddot{x}_j(t) + \cdots$$

symbolize a general linear dynamic relation between n variates $x_1, x_2, \cdots, x_n$; (instead of derivatives we might have finite differences or "lag" terms). We shall call (1) a *structural* dynamic equation *if* the variates and derivatives in the right-hand member may take on different values *independent of each other*, giving *alternatively* different values to x_i (e.g., (1) may be a dynamic demand function expressing the quantity demanded at a given point of time by alternative values of, say, price and income and their rates of change). The reason for the existence of such structural relations may be certain technical conditions, psychological laws, etc.

Let us now assume that we have $n-1$ other independent structural equations between the n variates in (1) ($i=1, 2, \cdots, n$), some of which may be static, others dynamic. And suppose these n equations give a simplified macrodynamic picture of the whole economic system. If these n equations are to hold simultaneously, the possible variation of the n variates is limited to the solutions of the system. Step by step we may then carry out elimination processes. During these processes new equations are derived from the original structural equations. Any such equation, the existence of which depends upon *two or more* structural equations, is called a *confluent relation.* The coefficients in confluent relations are composed of structural coefficients.[2] By the elimination process is built up a *hierarchic order* of confluent relations, at the top of which stand equations each containing only one time function x_i,

[1] The leading ideas in this study were originally given in a short paper read before the meeting of the Econometric Society at Oxford, 1936. (See report of the meeting, ECONOMETRICA, Vol. 5, Oct., 1937, pp. 373–374.) The author is greatly indebted to Professor R. Frisch, Oslo, and Professor J. Tinbergen, Geneva, for their valuable suggestions towards improving both the content and the form of the present paper.

[2] See for instance discussion between Professor Frisch and Dr. Marschak on this subject in the report of the meeting of the Econometric Society at Oxford 1936, ECONOMETRICA, Vol. 5, Oct., 1937, p. 374.

with a certain number of its derivatives or "lags." These last equations we call *final confluent relations.*

2. Theoretical Relations and Observed Confluency. A Special Formulation of the "Shock Theory"[3]

So far we have been speaking of the structural relations (1) as if they were *exact laws*, i.e., with absolutely fixed coefficients. If this were so, the time shape of the variates x_i observed in our system would be the exact solutions of (1) $(i=1, 2, \cdots, n)$. But this assumption is perfectly unrealistic. Even in physics one does not have such laws in the strict sense. *The structural equations must be taken as laws in the statistical sense, i.e., as average laws.* If individual observations (the set of observed values of the variates at individual points of time) are inserted in a theoretical structural equation, there will be a certain amount of "unexplained scatter." One could consider this as the result of fixing the structural coefficients at certain constant average values. If the structural equation is to be satisfied by *all* sets of observations the coefficients will have to vary. From this point of view each coefficient in (1) may then be written

$$(2) \qquad (\bar{a}_{ij} + \xi_{ij}^{(t)}), (\bar{b}_{ij} + \eta_{ij}^{(t)}), \cdots$$

where $\bar{a}$, $\bar{b}$, $\cdots$ are the constant average values of the coefficients and ξ, η, $\cdots$ are additional "error terms," which are functions of time. A set of coefficients such that:

(a) the error terms in different coefficients are independent;

(b) each error term is an erratic function of time, i.e., it has no systematic serial correlation (this is, of course, a very simplified condition);

(c) the error term exerts its influence on the coefficient only at the point of time for which the corresponding equation is supposed to hold; we shall call *stochastic structural coefficients* and the corresponding equations we shall call stochastic structural equations.

Now it is interesting to see what happens to a final confluent relation in such systems. The coefficients in the confluent relations are composed of stochastic coefficients from the system of structural relations. They may be written as a corresponding expression in the average values of the structural coefficients plus an error term. The distribution of this error term may be complicated. To have a clear picture of the

[3] Frisch, "Propagation Problems and Impulse Problems in Dynamic Economics," *Economic Essays in Honour of Gustav Cassel*, London, 1933; Frisch, "Circulation Planning," ECONOMETRICA, Vol. 2, July, 1934, p. 271; E. Slutzky, "The Summation of Random Causes as the Source of Cyclic Processes," ECONOMETRICA, Vol. 5, April, 1937, pp. 105–146.

problem let us consider a case where the final confluent equation is a second-order differ*ence* equation

$$x_i(t) = A_1 x_i(t-1) + A_2 x_i(t-2), \tag{3}$$

the coefficients A_1 and A_2 being for simplicity taken as products of certain stochastic structural coefficients. Then A_1 and A_2 may be written

$$A_1 = \bar{A}_1 + E_1^{(t)},$$
$$A_2 = \bar{A}_2 + E_2^{(t)},$$

where $\bar{A}_1$ and $\bar{A}_2$ are the corresponding products of average values of the structural coefficients, and $E_1^{(t)}$ and $E_2^{(t)}$ are error terms, which are functions of time with averages equal to zero. E_1 and E_2 may contain error terms for several points of time. They will usually not be independent, because the same structural coefficients may enter in both A_1 and A_2. Now the observations $x_i(t)$ fulfil exactly the equation (3), which also may be written

$$x_i(t) = \bar{A}_1 x_i(t-1) + \bar{A}_2 x_i(t-2) + E_1^{(t)} x_i(t-1) + E_2^{(t)} x_i(t-2). \tag{3*}$$

The equation (3*) is in fact similar to the nonhomogeneous difference equation of a function created by cumulation of "random shocks," the cumulation system (the weight system) being the solution of the homogeneous equation obtained by neglecting the error terms in (3*).[4] If the solution of (3*) with the error terms neglected is a damped sine curve, or one or two damped exponentials, the curve $x_i(t)$ as calculated directly from (3*) with error terms will be a more irregular curve showing some wave-like movements; i.e., it will have a shape more similar to *observed curves of economic variates. The random variation of the structural coefficients is sufficient* to give the curves this character. These phenomena are important in problems of statistical verification of a dynamic theory. Some discussion of these problems is given in II,3 below.

II. THE ANALYSIS OF PARTIAL DYNAMIC SYSTEMS

1. The Problem

In the foregoing chapter we have discussed how observed relationships of business-cycle variates depend upon an underlying structure of stochastic structural relations. In constructing this theoretical system of structural relations one has to make simplifications.[5] To make such

[4] Extensive theoretical and numerical studies on such equations have been directed by Professor Frisch at the University Institute of Economics, Oslo, Norway.

[5] An instructive review of different types of simplification is given by Professor Tinbergen in his article, "On the Theory of Business-Cycle Control," ECONOMETRICA, Vol. 6, Jan., 1938, pp. 29–33.

simplifications is a true art. If the system is to have any realistic character as a *total system*, there will be a certain number of important variates which we cannot neglect. In many cases we are, however, more interested in a detailed study of only a *certain group* of variates, without raising the whole business-cycle problem. In other words, it is a study of what may be called a *partial* macrodynamic system.

As an example, suppose we are interested in "explaining" the movements in one single variate x_i. Inspecting the total system of structural relations (1), we may happen to find a rather simple relation connecting this variate with, say, two other variates (or derivatives, etc.) x_h and x_k. If these two new variates are taken as given, the first one appears as being "explained." But this "explanation" is unsatisfactory. Looking more thoroughly around in the structural system, we shall find that the "explaining" variates appear in other structural relations as being "explained" by certain other variates, among which even x_i itself may occur. In this way the search for explanation is carried on automatically. If we wish to break off this search before having taken the whole system of structural relations into account, certain things have to be taken as *known*, in order to have a determinate system. The principle for choosing these known things must be such that the partial system *does not conflict with the true conditions in the total system*. This means, among other things, that we must not make our subsystem determinate by adding some more or less artificially invented structural relations. These may satisfy our craving for theoretical determinateness but may be unrealistic. They may not give a good approximate description of the conditions that prevail in the "large" total system. Another procedure would be to take some of the time series as given. But this usually would not permit a very profound analysis of the problem. There is a third procedure which seems more reasonable in many cases, namely the introduction of some confluent relations which we know exist.

2. Confluent Relations As a Means of Connecting a Partial System With the Total System

This principle can be formulated briefly as follows: Suppose we have a partial system of m variates between which there exist $k<m$ structural relations. Then the system is made determinate by introducing $m-k$ *confluent relations* between the m variates considered. None of these confluent relations must of course be derivable from the k structural relations and the other confluent relations considered. The introduction of these supplementary confluent relations is a way of letting the situation in the total system influence our study of the partial system. If we *had* worked with the total system, we could have

obtained the confluent relations considered by an elimination process. The point now is that *we do not carry through* this elimination process, but take the elementary confluent relations directly as they are *observed in statistical data.*

This procedure may be justified by the following considerations:

(a) The elimination process in the complicated total system may lead to certain *simple* confluent relations between a small number of variates. By consideration of the observed variates it is more easy to discover these simple connections than the system of complicated underlying structural relations. On the other hand, these simple observed facts are certainly *realistic* elements in the economic system. If we start with the construction of a total system of structural equations, these elementary confluent relations ought to come out as elimination results. Taking the elementary confluent relations as they are *observed* is therefore equivalent to having assumed an underlying *realistic* structure and to having carried out a *correct* elimination process.

(b) The compound coefficients in the elementary confluent relations may have relatively *smaller* error terms than the individual structural coefficients. This may be precisely the reason why they are easily discovered in the observations. This is of importance for the statistical verification as discussed in the section below.

(c) Taking certain elementary confluent relations as a datum is not so great a reduction in generality as taking the whole *time shape* of certain variates as given. The same confluent relation between several variates may indeed hold for different time shapes.

3. *The Problem of Statistical Verification*

The fundamental character of a dynamic system is defined by the average constant values of the structural coefficients. The main problem in statistical verification is therefore to estimate these coefficients. We have defined the structural coefficients as stochastic: they had to vary in order that the observations should fulfil the structural relations. Replacing these variable coefficients by their constant average values, we take the opposite viewpoint, saying: How can a set of constant coefficients be chosen so that the structural equation is satisfied "as well as possible," this expression being defined by some statistical principle, such as "least-squares deviations," etc. This raises the problem of multiple-regression analysis, certain special aspects of which must be discussed in this connection.

Suppose we wish to estimate the average coefficients in a final confluent relation in order to find the time shape. This is one way of statistical verification of theory. As the coefficients in a final confluent rela-

tion are composed of structural coefficients, one could try first to estimate these. For several reasons the latter procedure seems to be dangerous.

First, by definition the structural coefficients are coefficients of independent variates, while the variates observed are consequences of the whole system. If a theoretical structural relation is fitted, there is great danger of multicollinearity (or multidependencies if the relations are not linear) due to the other conditions which the system lays upon the variates. On the other hand, it may be difficult to see whether all the relations we have assumed structural really have this character, in particular to see if they are independent. In such cases, by fitting each structural equation separately and inserting the coefficients in the final equation, we should get a fictitious determinateness of the system.

Second, if an estimated structural coefficient is inserted in the final confluent relation, the error of estimation would reappear systematically in all those compound coefficients where it is involved. If certain elementary confluent relations whose error terms are small are accepted, this danger is reduced.

Third, some of the coefficients needed may belong to a structural equation involving variates for which it is impossible to have statistical observations.

When the problem is to find the time shape of a variate, it therefore seems better to fit the final confluent relation directly. In the simple case (3*) the condition for taking the elementary regression of $x_i(t)$ on $x_i(t-1)$ and $x_i(t-2)$ is that all unexplained scatter is found in this variate. This setting of the problem is unsatisfactory from our viewpoint. In fact, the unexplained scatter must be ascribed to the incompleteness with which the average economic laws *as such* describe the observed facts. For instance the equation (3*) could as well have been written (with an analogous definition of the coefficients)

$$\begin{aligned}(3^{**})\qquad &\bar{A}_0x_i(t) + \bar{A}_1x_i(t-1) + \bar{A}_2x_i(t-2)\\ &\qquad = E_0^{(t)}x_i(t) + E_1^{(t)}x_i(t-1) + E_2^{(t)}x_i(t-2)\end{aligned}$$

where there might be no systematic correlation between the error terms and any left-hand member. Each term in the average law—i.e., the expression to the left put equal to zero—will have to bear its part of this scatter. In general cases there will be systematic correlations between error terms and variables, and the magnitude of these correlations will usually be unknown. From this point of view it seems reasonable to choose some *mean* regression, for instance the "diagonal

regression."[6] This has been done in the following statistical applications.

III. A SIMPLIFIED PARTIAL SYSTEM IN THE DYNAMIC THEORY OF STOCK PRICES

As an illustration of the method of supplementary confluent relations we shall take a partial system of some important variates in the business-cycle movements: the movements of stock prices.

1. The Variates[7]

We define our variates in such a way that we may use only homogeneous equations, i.e., we take out possible constants beforehand. This simplification does not restrict generality. We further assume that such transformations can be made as will make our system linear. The following variates will come into consideration (subscript t indicates points of time):

X_t = stock prices (index),

(*) $x_t = \log X_t - \log X_0$ = logarithm of stock prices measured from their equilibrium position (if $X_0 = 1$ we have $x_t = \log X_t$),

r_t = stock dividend (per cent of nominal value),

ρ_t = general interest rate (the common bank rate) per cent,

(**) $\log r_t - \log \rho_t = \log (r_t/\rho_t)$ = logarithm of relative stock gain,

P_t = general price level,

(***) $q_t = \log P_t - \log P_0$ = logarithm of general price level measured from equilibrium position (if $P_0 = 1$, we have $q_t = \log P_t$).

Only those quantities indicated by asterisks will enter as unknown in our system, and the final purpose is to find the movement of x_t.

2. The System of Equations

In a stationary case one might consider the stock price as capitalised income from stocks, the capitalising factor being the rate of interest ρ_t; in other words,

$$\rho_t X_t = r_t \quad \text{or} \quad x_t = \log r_t - \log \rho_t. \tag{4}$$

In nonstationary cases there would be certain discrepancies between

[6] Frisch, *Statistical Confluence Analysis by Means of Complete Regression Systems*, Publication No. 5 of the University Institute of Economics, Oslo, 1934.

[7] O. Donner, "Die Kursbildung am Aktienmarkt," *Vierteljahrsheft zur Konjunkturforschung*, No. 36, Berlin, 1934. Many useful suggestions as to choice of variates and formulation of theory have been found in this book.

x_t observed and x_t calculated from (4). These discrepancies may contain a systematic part due to variations in the *risk element* of holding stocks. This risk element is attached to the movements in stock prices. In a simplified way this may be taken account of by changing (4) to

$$x_t = \log r_t - \log \rho_t + b\dot{x}_t, \tag{4*}$$

where b probably has a certain positive average value, but with a rather wide range of stochastic variation, due to the fact that the variate $\log (r_t/\rho_t)$ is a sensitive factor in the business cycle. Equation (4*) is a structural relation expressing a certain behavior of the stockholders. It contains only two unknowns, namely x_t and $\log (r_t/\rho_t)$.

For further "explanation" it is natural to introduce the new variate q_t (the logarithm of the price level). According to the Wicksellian theory the movements in price level are connected with the discrepancy between income from capital investment and the monetary rate of interest. As a very simplified expression for this relation we assume the equation

$$\log (r_t/\rho_t) = \mu\dot{q}_t + \nu q_t. \tag{5}$$

This equation indicates that the relative investment gain (represented by the relative stock gain) depends upon the rise or fall in the price level and on whether the price level is already high or low. Equation (5) may be considered as a confluent relation, since it is an elimination result of the structural relations which express how the capital investment depends upon technical productivity of capital and the capital gain.[8] Because of the very simple structural relations upon which this equation is built, the coefficients are probably not very stable.

Equation (5) gives one new equation and one new unknown variate q_t. There is therefore still one degree of freedom in our partial system. If we carry the search for "explanation" further, we shall soon get into the general problem of business-cycle movements. Instead of doing this, we now complete the system by introducing a well-known confluent relationship, namely the "lag" between the general price level and the stock prices, as we know it by observation. In other words we utilize the facts entering into the Harvard A-B-C-method.[9] The average lag may be determined (as is done in the Harvard studies) by the principle of maximum correlation. We then put

$$q_t = kx_{t-\theta}, \tag{6}$$

[8] See, for example, Professor Frisch's mimeographed lectures on monetary theory, with a discussion of the Wicksellian theory, Oslo, 1934.

[9] *The Review of Economic Statistics*, Harvard University Committee on Economic Research, Cambridge, Mass., Vol. 1, 1919.

where k and θ must be considered as stochastic coefficients having certain positive average values.

The equation (6) is certainly in the mathematical sense independent of our two first equations. If these hypotheses are accepted as they stand, the system is therefore determinate.

From the equations (4*), (5), and (6) we easily derive the following final confluent relation:

$$x_t - b\dot{x}_t - k\nu x_{t-\theta} - k\mu\dot{x}_{t-\theta} = 0. \tag{7}$$

3. Special Cases Leading to a Simplification of the Final Equation

In the final equation (7) certain interesting special cases may occur, by the vanishing of certain coefficients. First it may be noticed that it would be against all observations to let k vanish; k must be assumed to have a significant positive average. We are then left with the coefficients b, μ, ν. From a theoretical point of view one could hardly imagine that the capital gain should be independent of the price level and its movements. Therefore, the case of vanishing μ and ν seems of little interest. As to the coefficient b, the situation is different. During "unquiet" periods the element $b\dot{x}_t$ will probably play a rather important role, while under "quiet" periods it may perhaps be neglected. This neglect may be justified by the fact that the variate $\log(r_t/\rho_t)$ is a sensitive business-cycle element, so that most of the variations in x_t may be "explained" by this element alone. Putting b equal to zero, we get from (7)

$$\dot{x}_t = ax_t - cx_{t-\theta'}, \qquad \begin{cases} a = -\nu/\mu, \\ c = -1/k\mu, \\ \theta' = -\theta. \end{cases} \tag{8}$$

This is a well-known type of dynamic equations.[10] A general discussion of the equation and its numerical solution is given by Frisch and Holme.[11]

From (7) is seen that an equation of the type (8) also appears when μ is zero. It is important to keep this in mind during the statistical verification. Indeed if the neglect of the term $b\dot{x}_t$ gives a good fit, this will also be obtained by neglecting alternatively the term $k\mu\dot{x}_{t-\theta}$, because we then fit the same type of equation. Examples will be found in the statistical applications below. The fact that (8) fits the observa-

[10] See, for example, M. Kalecki, "A Macrodynamic Theory of Business Cycles," Econometrica, Vol. 3, July, 1935, pp. 327–344.

[11] "The Characteristic Solutions of a Mixed Difference and Differential Equation Occurring in Economic Dynamics," Econometrica, Vol. 3, April, 1935, pp. 225–239.

tions is thus not sufficient for putting μ equal to zero, if it cannot be justified by theoretical considerations.

4. Note on the Numerical Solution of the Final Equation (7)

The equation (7) may be written

$$(7^*)\qquad x_t + a_1\dot{x}_t + a_2x_{t-\theta} + a_3\dot{x}_{t-\theta} = 0, \qquad \begin{cases} a_1 = -b, \\ a_2 = -k\nu, \\ a_3 = -k\mu. \end{cases}$$

Putting as usual

$$(9)\qquad x_t = Ce^{\kappa t},$$

we obtain the characteristic equation

$$(10)\qquad e^{-\kappa\theta} = -\frac{1 + a_1\kappa}{a_2 + a_3\kappa}\cdot$$

By the transformations

$$(11)\qquad u = \frac{\theta}{a_3}(a_2 + a_3\kappa),$$

$$(12)\qquad H = -\frac{a_1}{a_3}e^{-\theta(a_2/a_3)} = \text{constant},$$

$$(13)\qquad K = \theta\frac{a_1a_2 - a_3}{a_3{}^2}\cdot e^{-\theta(a_2/a_3)} = \text{constant},$$

the equation (10) takes the simple form:

$$(10^*)\qquad ue^{-u} = Hu + K.$$

From this equation the real solutions may be determined graphically as the intersections of the curve $y = ue^{-u}$ and the straight line $y = Hu + K$; see Figure 1. To each value of u there corresponds one and only one value of κ; it is therefore sufficient to discuss the u-solutions. From $y = ue^{-u}$ it follows that:

$$(14)\qquad \frac{dy}{du} = (1 - u)e^{-u},$$

$$(15)\qquad \frac{d^2y}{du^2} = (u - 2)e^{-u}.$$

As e^{-u} is always positive, $u = 1$ will be a maximum and the only one. Furthermore there is one and only one point of inflection, namely $u = 2$. From this follows that the straight line may have 0, 1, 2, or a maximum of 3 intersections with the curve $y = ue^{-u}$.

To find the complex roots we put

$$(16)\qquad u = \gamma + i\delta, \qquad e^{-u} = e^{-\gamma}(\cos\delta - i\sin\delta), \qquad i = \sqrt{-1},$$

where γ and δ are real. From (10*) we then obtain the two conditions

$$e^{-\gamma}\cos\delta = H + \frac{K\gamma}{\gamma^2+\delta^2}, \tag{17}$$

$$e^{-\gamma}\sin\delta = \frac{K\delta}{\gamma^2+\delta^2}\,. \tag{18}$$

Dividing (18) by (17) we get

$$\tan\delta = \frac{K\delta}{H(\gamma^2+\delta^2)+K\gamma}$$

$$\gamma = -\,K/2H \pm \tfrac{1}{2}H\sqrt{K^2+4HK(\delta/\tan\delta)-4H^2\delta^2}. \tag{19}$$

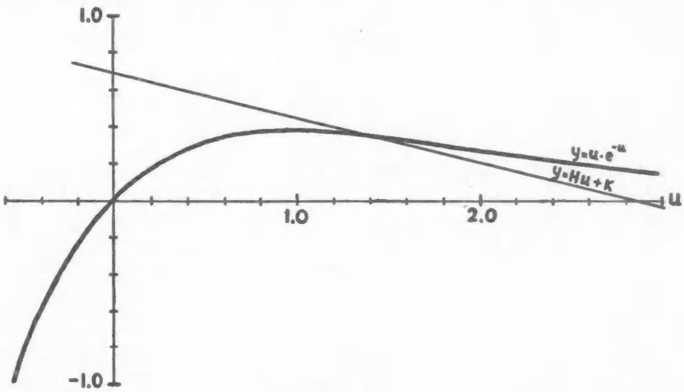

FIGURE 1

From this expression γ may be plotted as a function of δ. By inserting points along this curve in one of the equations (17) or (18) it will usually be easy to approximate the solutions.

IV. STATISTICAL APPLICATIONS

We shall give some results obtained by fitting the final equation (7) to actual data. This is essentially a matter of multiple-regression analysis. For the fitting we have built entirely on the "bunch technique" developed by Professor Frisch.[12] The validity of the results obtained is judged by inspection of the "bunch maps."

1. Norwegian Data

A monthly index of stock prices ("all stocks") is published by the

[12] See footnote 6.

Statistical Central Bureau, Oslo.[13] The series is computed as an average of buyers' prices in per cent of par, from the quotations of the Oslo Börs (the Stock Exchange). We have chosen the postwar period, 1920–26, in order to have a period with great variations, and to have a homogeneous series (in certain later years the calculation technique for the index is changed). To simplify the computation work by the regression analysis only the months March, June, September, and

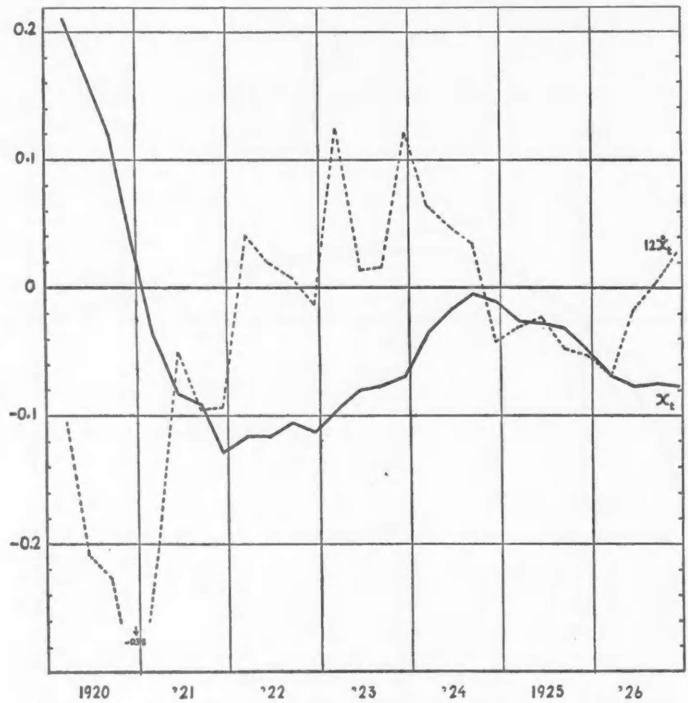

FIGURE 2.—Stock Prices, Norway 1920–26. x_t and $12\dot{x}_t=(x_{t+3}+x_{t+2}+x_{t+1}-x_{t-1}-x_{t-2}-x_{t-3})$, plotted quarterly. x_t: ——, $12\dot{x}_t$: - - - -.

December are used, but t below is always measured in units of one month. The series is divided by 100 and X_0 is put equal to 1. Then x_t is equal to $\log X_t$. No smoothing is applied. The derivative $\dot{x}_t$ is approximated by the linear operation

$$\dot{x}_t = (x_{t+3} + x_{t+2} + x_{t+1} - x_{t-1} - x_{t-2} - x_{t-3})/12. \tag{20}$$

[13] *Statistiske Meddelelser,* Monthly Bulletin of the Statistical Central Bureau, Oslo.

The graphs of x_t and $\dot{x}_t$ are given in Fig. 2.

The lag θ is estimated to be 6 months, from a graphical comparison between x_t and q_t (the logarithm of wholesale price index as published by the Statistical Central Bureau).[14]

In the regression analysis we have used origin correlations, i.e., no average is subtracted from the data, as the "normal level" is assumed to be removed by log X_0.

The result of the complete regression analysis may be studied in the bunch map given in Figure 3. The regression variates are numbered from 1 to 4 in the sequence in which they stand in equation (7).

The first remarkable result is the good fit in the sets (123) and (134), corresponding to an equation of the form (8). Taking the diagonal regression we obtain

$$x_t - 6.71\dot{x}_t - 1.12x_{t-6} = 0 \quad \text{[in the set (123)]} \tag{21}$$

and

$$x_t - 0.87x_{t-6} - 5.08\dot{x}_{t-6} = 0 \quad \text{[in the set (134)].} \tag{22}$$

Transforming these equations into the form (8) and solving by means of the Frisch-Holme method[15] we obtain the following results:

(23) In the set (123): A damped sine curve of about 82 months' period, with a damping coefficient of $e^{-0.005}$ or about $e^{-0.4}$ per period.

(24) In the set (134): A damped sine curve of about 74 months' period, with a damping coefficient of $e^{-0.018}$ or about $e^{-1.4}$ per period.

(There exist further solutions giving cycles shorter than one year).

The possibility exists that such results are fictitious, created by ourselves as an effect of manipulations with the erratic element in the series. Then the results will depend on the method of fitting, and not on the data. This is certainly not the case in the present results, as may be easily verified by constructing the theoretical moment matrix obtained by applying our method of fitting to random numbers. It turns out to be a diagonal matrix and cannot explain the good fit in our bunch map. Similar tests are made for the other results below. Further, the period of 6–7 years may clearly be *seen* in the graph of x_t (Figure 2).

The equations (21) and (22) give certainly a good *description* of the x_t-curve, as is shown by the tight bunches for the coefficients. From the theoretical point of view it is, however, perhaps still more interesting that we may pass on to the total four-set of regression variates without losing very much in bunch tightness. The diagonal regression gives here the following result

[14] *Ibid.*

[15] See footnote 11.

$$x_t - 4.38\dot{x}_t - 1.02x_{t-6} - 4.04\dot{x}_{t-6} = 0. \tag{23}$$

Applying the method of numerical solution described in III, 4, we obtain first one (and only one) real root of the characteristic equation, namely

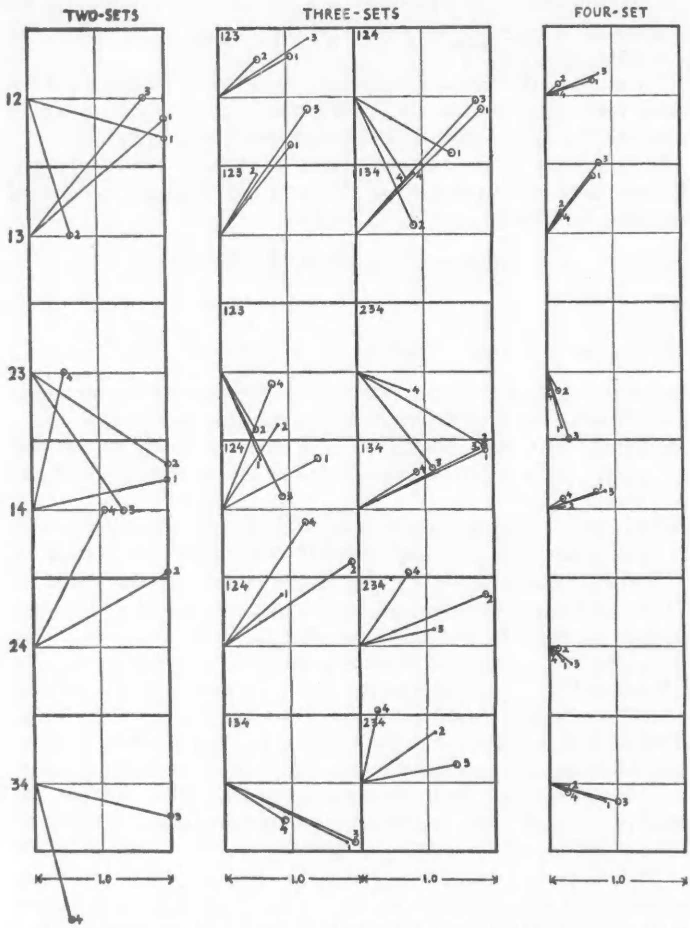

FIGURE 3.—Bunch Map, Norwegian Stock Prices. Variate No. 1: x_t, No. 2: $\dot{x}_t$, No. 3: x_{t-6}, No. 4: $\dot{x}_{t-6}$.

(24) a damped exponential with damping factor about $e^{-0.01}$, or $e^{-0.12}$ per year.

For the complex roots we obtain as main solution

(25) a short damped cycle of about 21 months with damping factor about $e^{-0.015}$.

There further exist shorter, but no longer, cycles. Some remarks on this result are given in connection with the American data below.

From (7) is seen that we here have the special case of a structural coefficient, b, *entering isolated in the final confluent relation.* From (23) we get $b=4.38$. This may perhaps be taken as an estimate of b in (4*). It means that if, e.g., $\dot{x}_t=0.01$—that is, the stock price increases by $0.01/\log_{10}e$ or about 2.3 per cent per month—then the stock price will lie about 10 per cent higher than the capitalized value of stock dividends, which seems rather realistic. One could perhaps interpret this as if the stockholders look ahead roughly $10/2.3=$about $4\frac{1}{2}$ months, based upon today's stock price movements.

2. American Data

For further application of the theory we have chosen the Harvard series: Price of 12 industrial stocks, 1903–14 (*Wall Street Journal*, Dow-Jones and Co.).[16] The series is given as deviations from trend measured in units of standard deviations. For our purpose we have transformed this into total ordinate divided by trend. The logarithm of this series is given in Figure 4. The series $\dot{x}_t$ is here simply determined by

$$\dot{x}_t = \tfrac{1}{2}(x_{t+1} - x_{t-1}), \tag{26}$$

because the movements in the present series appear to be shorter than in the Norwegian series. Figure 4 also shows $\dot{x}_t$. The lag θ is put equal to 9 months.[17] The selection of months and the regression technique are exactly the same as for the Norwegian data.

The bunch map of these data (not given here) turned out quite different from that of the Norwegian series. The bunches were mostly quite open or "exploded." Only one set showed a rather good fit, namely the set (134) corresponding to an equation of the form (8). The diagonal regression calculated in this set is

$$x_t - 0.83x_{t-9} - 4.15\dot{x}_{t-9} = 0, \tag{27}$$

and these coefficients are almost *unaltered* by adding the variate No. 2 ($\dot{x}_t$), i.e., in the total four-set (this set is "exploded" for the rest of the coefficients).

This result may be a verification of our hypothesis in Chapter III about the coefficient b. The period before the war is usually mentioned

[16] *Review of Economic Statistics*, Vol. 1, 1919, p. 191, Table M.
[17] *Ibid.*, p. 182.

as an example of a "quiet" and "normal" period, which may explain the nonsignificance of a constant value of the coefficient b. The characteristic equation of (27) gives as principal solution

(28) Two damped exponentials with damping factors $e^{-0.06}$ and $e^{-0.11}$ (per month),

i.e., the simplified equation (27), which here may be theoretically justified, gives the *same type of principal solution as* the complete equation in the Norwegian data. This seems to indicate that damped exponentials are a characteristic type of solutions in this field.[18] This

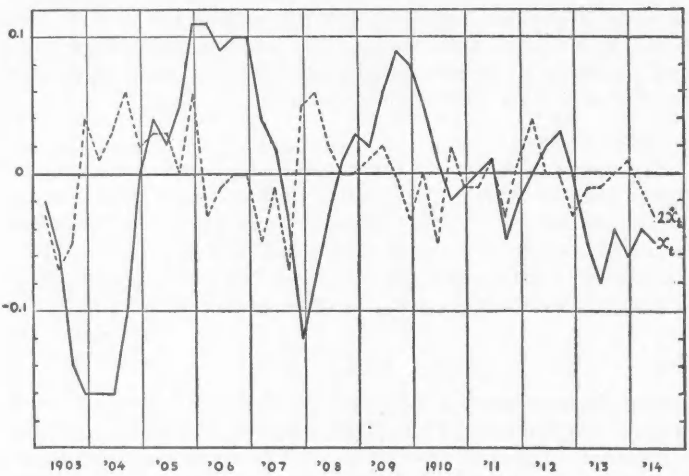

FIGURE 4.—Stock Prices, U.S.A. 1903–14. x_t and $2\dot{x}_t=(x_{t+1}-x_{t-1})$, plotted quarterly. x_t: ——, $2\dot{x}_t$: - - - - .

is not in contradiction to the *observed cyclical* character of the series, which, as explained in I, 2, may be the result of the stochastic variation of the coefficients. Because of this variation, the corresponding characteristic solutions of the final equation taken on an instantaneous basis may even oscillate between exponentials and cycles. The sprawling in the bunch map of the coefficients shows that there is room for such variations.

University Institute of Economics,
Oslo

[18] See "Editor's note" in ECONOMETRICA, Vol. 6, January, 1938, p. 31.

ANNUAL SURVEY OF ECONOMIC THEORY: THE THEORY OF DEPRECIATION

By Gabriel A. D. Preinreich

The literature of depreciation is very extensive, but it clearly shows that the study of this important subject is being carried on in separate compartments isolated from one another. It is the purpose of the present article to appraise representative samples of independent thought and to expand their valid portions into rudiments of a theory, or a foundation for further development.

Mathematical economists have excogitated more or less fine-spun applications of the law of capital value to so-called capital assets, but most of their attention appears to have been devoted to the economic behavior of a single "machine," rather than to the continuous flow of productive assets through a plant or production center. There has been great interest in why the economic life or usefulness of a machine reaches its end, but comparatively little in the mass phenomena of how gradual consumption actually occurs.

Engineers, on the contrary, emphasize the latter aspect in elaborate mortality studies conducted upon the same principles as for human lives. The results are expressed in time-service units, often considered equivalent to cost units. To describe these units as life units would be more accurate, since the capacity or opportunity for rendering service within successive time units is generally variable. Little or no attention is given to the concept of value, as governed by the economic phenomena of profit and interest. The tendency throughout is to subordinate theory to practice. Thus, for instance, entirely unrelated types of curves are commonly fitted to groups or phases of data which are of the same type, in the sense that one group bears a definite and fixed mathematical relationship to the other.

Accountants are even more "practical-minded" than engineers, and this possibly explains the lack of any notable contribution on their part. In publications, they are usually content to restate what has been said before. The various depreciation methods employed in practice are always described in the same terms, with hardly any attempt at scientific analysis. Nevertheless, some of the truisms and generalities recopied and re-expounded from year to year have more merit than economists are willing to concede.

Business men and management experts have also contributed millions of words to the theory of depreciation, it being quite apparent in many cases that the wish is the father of the thought. This is true particularly in the public-utility field, where the battle still rages with

unabated vigor on the question of whether or not a used machine can be worth less than it cost, so long as it serves as well as a new one.

I

To emphasize that the stream of machines gradually consumed and replaced is more important than its component elements, it seems best to begin with a review of engineering studies of mortality. Dr. Edwin B. Kurtz[1] has compiled a wealth of actual data on 52 different kinds of equipment, all of which show a striking similarity of behavior.

When a large number of similar items of equipment is installed at the same time, the rate at which they will drop out of service forms a bell-shaped, but usually skew, *frequency distribution*. Let it be drawn in such a form that its ordinates $f(0)$ at the time $t=0$ and $f(n)$ at the time $t=n$ be zero, and that the total area enclosed by the curve and the axis of abscissas be equal to unity. For any other time t the curve will be denoted by $f(t)$. Summation of $f(t)$ from t to n gives the *mortality curve* per centum of the total number of machines installed at the time $t=0$:

$$(1) \qquad M(t) = \int_t^n f(\tau)d\tau; \qquad M(0) = 1; \qquad M(n) = 0.$$

The area enclosed by the mortality curve and the co-ordinate axes represents the total life units of service expected from the machines installed at the outset. Since the ordinates are expressed per centum, the area also equals the *average life* of the machines, i.e., their life expectancy at $t=0$:

$$(2) \qquad a = \int_0^n M(\tau)d\tau.$$

The *life expectancy* of the machines in service at any other time t is the quotient of the remaining life units of service by the mortality curve:

$$(3) \qquad L(t) = \frac{1}{M(t)}\int_t^n M(\tau)d\tau.$$

The sum of the age t and the life expectancy $L(t)$ is the *probable life* of any machine still in service. If the rate of service rendered by each machine is contant, the remaining life units of service, expressed per centum of the original units, may be called the *remaining services per centum*. For a single machine, the information is obtained by dividing

[1] *Life Expectancy of Physical Property*, New York, Ronald Press Co., 1930.

the life expectancy by the probable life. In terms of all machines originally installed, the formula is

$$R(t) = \frac{1}{a}\int_t^n M(\tau)d\tau. \tag{4}$$

From the analysis of the life characteristics of the machines, it is possible to compute the *rate of renewals* required to keep the number of machines in service constant. The rate of replacement of the original machines is evidently the basic frequency distribution. In addition, however, the replacements must also be replaced, it being assumed that the new machines will behave in the same way as the original ones.

Let us divide the area of the frequency distribution $f(t)$ and that of the mortality curve $M(t)$ into vertical strips of equal width Δt, beginning at the origin and proceeding toward the right. If k be the total number of strips, $t = k\Delta t$. The area of the first strip $f_1\Delta t$ represents the number of machines (per centum) which must be left at the end of the period Δt_1, out of the unknown number $u_1\Delta t$ actually added during Δt_1. Similarly $f_1\Delta t + f_2\Delta t$ is the number which must be left at the end of $2\Delta t$, out of the unknown total $u_1\Delta t + u_2\Delta t$ added during $2\Delta t$, and so on up to $k\Delta t$. But the average height of each successive strip of the mortality curve, taken from t backward, expresses that proportion of each successive addition, which is still in service at the time t. Thus, $M_k u_1 \Delta t$ is left of the machines installed during Δt_1, $M_{k-1}u_2\Delta t$ of those installed during Δt_2, and so forth. Therefore

$$\sum_{i=1}^{k} f_i\Delta t = \sum_{i=1}^{k} M_{k+1-i}u_i\Delta t, \qquad i = 1, 2, 3, \cdots, k \tag{5}$$

and

$$u_k\Delta t = \frac{\sum_{i=1}^{k} f_i\Delta t - \sum_{i=1}^{k-1} M_{k+1-i}u_i\Delta t}{M_1}. \tag{6}$$

By successively placing $k = 1, 2, 3, \cdots$, the total renewal curve may be constructed step by step. That is no doubt the manner in which Dr. Kurtz has computed his table 50 for the total annual renewals of property group VII.[2]

The *equation of the renewal curve* does not appear to have been developed by anyone up to the present. It can be found by taking the limit of (6). For $\Delta t = dt = 0$, $M_1 = 1$, and

$$0 = \int_0^t f(\tau)d\tau - \int_0^t u(\tau)M(t-\tau)d\tau. \tag{7}$$

[2] His own explanation is somewhat more complicated. Cf. *op. cit.*, pp. 168–176.

This is a *Volterra equation of the first kind.* Its derivative,

$$u(t) = f(t) + \int_0^t u(\tau)f(t - \tau)d\tau, \tag{8}$$

is known as a *Volterra equation of the second kind.* The literature of the latter is extensive, although it is largely restricted to linear forms. No specific reference to frequency distributions is made anywhere; nevertheless some of the methods discovered are applicable.[3] Limiting comment to the present problem, it may be said that, whenever it is possible to transform (8) into a differential equation of finite order, that seems the best way of solution. This can often be accomplished either by differentiating until the kernel $f(t-\tau)$ vanishes or—if it is indefinitely differentiable—until previous derivatives can be so transformed or combined that the integral will be the same as in the last derivative. If one equation is then subtracted from the other, the integral vanishes.

When this method fails, a solution in the form of an infinite series is still possible by expanding both sides, performing the integration term by term, and equating the coefficients of equal powers of t. That is really what most of the other methods amount to, even in cases where a differential equation of finite order could have been obtained. They merely afford various short cuts and, in simple cases, lead to the summation of the resultant series.

Dr. Kurtz has computed the value of the Pearsonian criterion K for each of his seven property groups and finds it negative throughout.[4] From this he concludes that the type of frequency distribution to be fitted to his data is in all cases the Pearsonian curve I, viz.:

$$y = y_0\left(1 + \frac{t}{a_1}\right)^{m_1}\left(1 - \frac{t}{a_2}\right)^{m_2}. \tag{9}$$

Differences in the behavior of the seven property groups are expressed by variations in the values of the constants. By way of illustration, let us place $m_1 = 1$, $m_2 = 2$, and $a_1 + a_2 = n$. The result closely reflects the behavior of group VII, which happens to be the largest group, embracing 17 different types of industrial equipment out of the 52 examined.

When the start of the curve is transposed to the origin and the area of the bell is reduced to unity, equation (9) takes the form

$$f(t) = \frac{12}{n^4}\, t(n - t)^2. \tag{10}$$

[3] For a survey of these methods and an exhaustive list of references see: *Indiana University Studies* by H. T. Davis, especially Nos. 88–90.

[4] *Op. cit.*, p. 91.

If the frequency distribution (10) is substituted in the general formula (8), four successive differentiations lead to a differential equation of the fourth order, the solution of which can be written at sight:

$$u(t) = Ae^{\alpha t} + Be^{\beta t} + e^{\gamma t}(C \cos \delta t + D \sin \delta t). \tag{11}$$

The roots α, β, and $\gamma \pm \delta i$ are found by solving the quartic auxiliary equation, while the constants A, B, C, and D are obtainable from four simultaneous linear equations expressing relationships found in the process of differentiating both (8) and (11). It should be noted that equation (11) is valid only for the first life cycle, $0 \leqq t \leqq n$, because the curve $f(t)$ beyond the bell must be eliminated. For each subsequent cycle a new equation is needed in the general form

$$u_{j+1}(t) = \int_{t-n}^{nj} u_j(\tau)f(t-\tau)d\tau + \int_{nj}^{t} u_{j+1}(\tau)f(t-\tau)d\tau, \qquad j = 1, 2, 3, \cdots; \tag{12}$$

where $u_j(t)$ is known from the previous calculation. It is unnecessary to continue beyond $j=2$, because seven-place logarithms soon become unable to distinguish the result from the straight line $1/a$. The solution of $u_j(t)$ is similar to (11), except that the coefficients are not constants, but polynomials of degree $j-1$ in t.

The renewal curve will consist of separate equations pieced together, whenever the bell of the frequency distribution has a finite range. Where the range is infinite, there will be only a single equation, valid until $t=\infty$. In either case, the general form of the curve resembles the damped oscillations of the governors of steam turbines, etc., when they seek a new equilibrium. The resemblance is closer when the bell has an infinite range, because it is then essential that the real part of the complex roots of the auxiliary equation be negative. Damping can not occur otherwise.[5] When the frequency distribution (10) is employed, the real part of the complex roots is positive and damping takes place merely because the equation is changed. This may be of interest to Dr. Kurtz, who considers the renewal curve as a simple damped sine function, as soon as it crosses its axis $1/a$ for the first time. Judging by his own data,[6] such a curve appears to fit none too well, especially for property group VII. When the real equation of the renewal curve is very complicated, (6) furnishes a more satisfactory approximation, provided Δt be taken small enough. Another simplification, which Dr. Kurtz adopts, is the use of the Gompertz-Makeham formula in lieu of

[5] For requirements in the construction of governors, which have a general bearing on the renewal curve, see W. Hort, *Technische Schwingungslehre*, Berlin, Springer, 1922, pp. 266 *et seq.*

[6] Kurtz, *op. cit.*, pp. 178–181.

the actual mortality curve (1). Such steps are often necessary in practice, but it is also worth bearing in mind what the true relationships are.

II

Attempts to compare various methods of depreciation in terms of a composite plant were greatly hampered up to the present by the lack of a simple general expression for the rate of renewals. With the aid a continuous renewal function[7] the task becomes easy.

To make the conclusions as general as possible, *expansion and increasing replacement costs* will also be introduced. It follows immediately from equation (7) that, if the plant is not only maintained, but also expanded at any variable rate $x(t)$, the curve of replacements and additions is obtainable from the relation

$$\int_0^t f(\tau)d\tau + e^{\int_0^t x(\tau)d\tau} - 1 = \int_0^t u(x, \tau)M(t - \tau)d\tau; \tag{13}$$

$$u(x, t) = f(t) + x(t)e^{\int_0^t x(\tau)d\tau} + \int_0^t u(x, \tau)f(t - \tau)d\tau. \tag{14}$$

When the price of new machines increases at any variable rate $q(t)$, the renewal function in terms of money may be written for brevity

$$u(x, q, t) = u(x, t)e^{\int_0^t q(\tau)d\tau}. \tag{15}$$

No matter what method of depreciation be employed, the *book value $C(x, q, t)$ of a composite plant* consists of the book value $c(t)$ of the original machines and that of all renewals and additions:

$$C(x, q, t) = c(t) + \int_0^t u(x, q, \tau)c(t - \tau)d\tau \qquad \text{for } t < n \tag{16}$$

and

$$C(x, q, t) = \int_{t-n}^t u(x, q, \tau)c(t - \tau)d\tau \qquad \text{for } t > n. \tag{17}$$

The book value $c(t)$ includes both services not yet rendered and the scrap value. If the scrap value of a single machine be $s(t)$, that of all original machines in service at the time t is $\int_t^n s(\tau)f(\tau)d\tau$. These machines depreciate at the rate of change of the difference between the book value and the scrap value. An operation similar to that indicated in (16) leads to the result:

[7] For purposes of general discussion, the symbol u will represent the renewal function for the entire interval $0 \leqq t \leqq \infty$, irrespective of how many separate equations it may actually consist of.

$$(18) \qquad D(x,q,t) = C'(x,q,t) - S'(x,q,t) - [1 - \int_0^n s(\tau)f(\tau)d\tau]u(x,q,t).$$

The *composite rate of depreciation* is obtained by finding the rate of change of the composite book value and deleting from the result both the rate of change of the composite scrap value and the renewals of the services consumed.

The best-known methods of depreciation may now be compared by developing the book-value curve $c(t)$ of the original machines, as defined by each. The methods selected are: 1. the straight-line method; 2. the annuity method (equivalent to the sinking-fund method when only one rate of interest is used); 3. the diminishing-balance method; and 4. the retirement method (here called a depreciation method purely for convenience of language). All methods assume that the scrap value of a single machine is a constant proportion of its purchase price.

The *straight-line method* is directly connected with mortality theory in the sense that it counts the unexpired life units and considers them equivalent to the service or value units to be recovered. To these units the scrap value must be added. The book value of many machines installed at the same time is therefore by (4):

$$(19) \qquad c_s(t) = \int_t^n \left[\frac{b}{a} M(\tau) + sf(\tau)\right] d\tau; \qquad b = 1 - s.$$

The *annuity method* makes the same assumptions, but it also discounts the value of both future services and scrap proceeds at some rate of return i. This procedure changes (19) to:

$$(20) \quad c_a(t) = \int_t^n [\phi M(\tau) + sf(\tau)]e^{i(t-\tau)}d\tau; \quad \phi = \frac{1 - s\int_0^n f(\tau)e^{-i\tau}d\tau}{\int_0^n M(\tau)e^{-i\tau}d\tau}.$$

The *diminishing-balance method* is a crude attempt to recognize a decline in the value of services, as a machine grows older. Instead of applying the constant depreciation rate b/a to the original cost of the machines still in service, it applies the constant rate $k = \frac{1}{a}\log_e s$ to the diminishing book value:

$$(21) \quad c_d(t) = -\frac{1}{k}[c_d'(t) + sf(t)] = e^{-kt} - s\int_0^t f(\tau)e^{k(\tau-t)}d\tau.$$

The *retirement method* is based upon the false claim that a machine which serves as well as a new one can not be worth less than it cost.

Depreciation is disregarded altogether, until the machine is scrapped. At that time the entire difference between cost and scrap value becomes an operating expense. The book-value curve of the machines originally installed accordingly equals the mortality curve:

$$c_r(t) = M(t). \tag{22}$$

Formulae (19) to (22) must now be inserted in the general expression (16) to find the composite book value for all four methods. Special cases lead to simple answers. For instance when replacement costs are static and expansion proceeds at a constant rate, the derivative of (16) is, for the straight-line method:

$$C_s'(x, t) = -\frac{b}{a} e^{xt} - s[u(x, t) - xe^{xt}] + u(x, t). \tag{23}$$

The reduction is made possible by the relations (13) and (14). The simplified composite-book-value formula is then:

$$C_s(x, t) = 1 + \int_0^t \left[bu(x, \tau) + \left(sx - \frac{b}{a} \right) e^{x\tau} \right] d\tau. \tag{24}$$

Proceeding similarly for the other methods, we obtain

$$C_a(x, t) = e^{it} + \int_0^t [bu(x, \tau) + (sx - \phi)e^{x\tau}]e^{i(t-\tau)} d\tau, \tag{25}$$

$$C_d(x, t) = e^{-kt} + \int_0^t [bu(x, \tau) + sxe^{x\tau}]e^{k(\tau-t)} d\tau \tag{26}$$

$$C_r(x, t) = e^{xt}. \tag{27}$$

The corresponding depreciation curves are by (18):

$$D_s(x, t) = -\frac{b}{a} e^{xt}, \tag{28}$$

$$D_a(x, t) = -\phi e^{xt} + iC_a(x, t), \tag{29}$$

$$D_d(x, t) = -kC_d(x, t), \tag{30}$$

$$D_r(x, t) = -b[u(x, t) - xe^{xt}]. \tag{31}$$

So far all curves are given in terms of the original cost of the plant, which was the unit chosen for the purpose. Further study is facilitated by expressing results in terms of the expanding cost of all machines in service at any time t. This base is the book value (27) obtained by the retirement method. If all formulae are divided by e^{xt}, the curves will in due course become horizontal lines. Their respective levels are of

interest. To find these ultimate levels, that of the rate of replacements and additions must be found first. For $t=\infty$ and $x(t)=x$, equation (13) may be written:

$$e^{x\infty} = u(x, \infty)e^{-x\infty}\int_{\infty-n}^{\infty} M(\infty - \tau)e^{x\tau}d\tau = u(x, \infty)\int_0^n M(\tau)e^{-x\tau}d\tau,$$

$$(32) \qquad u(x, \infty)e^{-x\infty} = \frac{1}{\int_0^n M(\tau)e^{-x\tau}d\tau} = \chi.$$

The ultimate discounted levels of curves (24) to (31) are now given by the necessary condition:

$$(33) \quad \lim_{t=\infty} \frac{d}{dt}\left[C(x, t)e^{-xt}\right] = 0; \quad C(x, \infty)e^{-x\infty} = \frac{1}{x}C'(x, \infty)e^{-x\infty}.$$

Therefore

$$(34) \qquad C_s(x, \infty)e^{-x\infty} = \frac{1}{x}\left(b\chi + sx - \frac{b}{a}\right),$$

$$(35) \qquad C_a(x, \infty)e^{-x\infty} = \frac{1}{x-i}(b\chi + sx - \phi),$$

$$(36) \qquad C_d(x, \infty)e^{-x\infty} = \frac{1}{x+k}(b\chi + sx),$$

$$(37) \qquad C_r(x, \infty)e^{-x\infty} = 1,$$

$$(38) \qquad D_s(x, \infty)e^{-x\infty} = -\frac{b}{a},$$

$$(39) \qquad D_a(x, \infty)e^{-x\infty} = -\phi + iC_a(x, \infty)e^{-x\infty},$$

$$(40) \qquad D_d(x, \infty)e^{-x\infty} = -kC_d(x, \infty)e^{-x\infty},$$

$$(41) \qquad D_r(x, \infty)e^{-x\infty} = -b[\chi - x].$$

Formulae for entirely static conditions are readily obtained by placing $x=0$, in which case $\chi=1/a$. A slight difficulty arises only for the straight-line method, since equation (34) becomes indeterminate and must be evaluated.

$$(42) \quad C_s(\infty) = \frac{0}{0} = \lim_{i=0} C_a(\infty) = s + \frac{b}{a^2}\int_0^n\int_\tau^n M(\nu)d\nu d\tau.$$

When replacement costs are changing, the general equation (16) can not be simplified to any considerable extent. A simple case was pre-

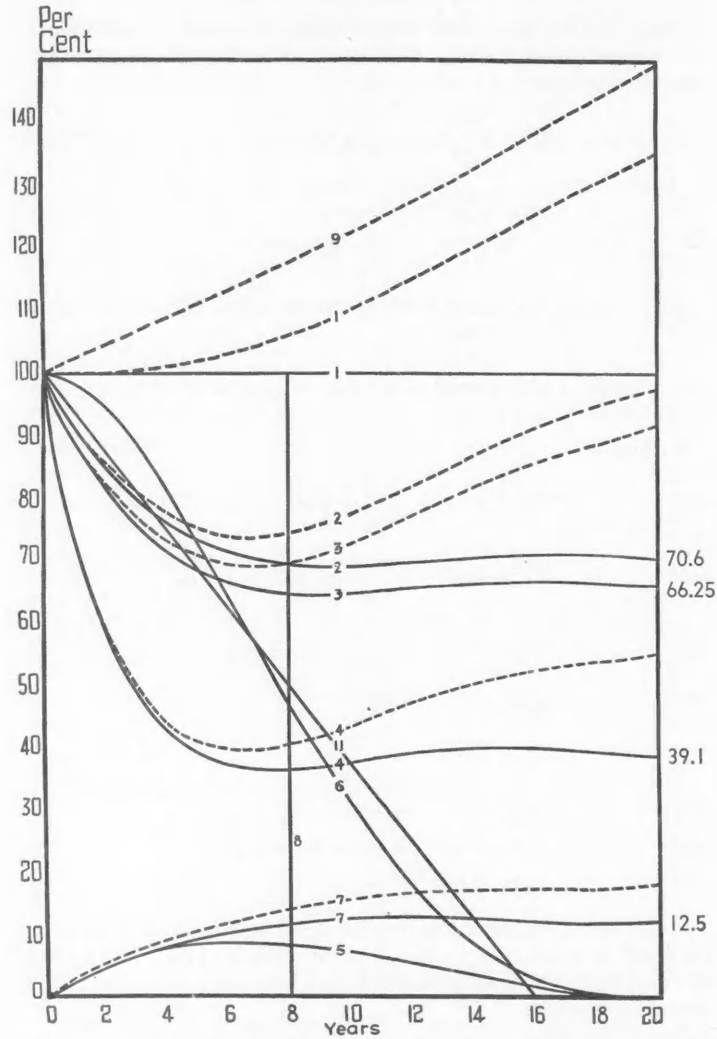

FIGURE 1.—Vertical scale represents per cent of cost of original plant. Smooth lines indicate static conditions. Broken lines include increase in replacement costs at the force of 2 per cent per annum.

Composite-book-value curves: (1) Retirement method; (2) Annuity method at the force of 7 per cent per annum; (3) Straight-line method; (4) Diminishing-balance method.

Auxiliary curves: (5) Basic frequency distribution; (6) Mortality curve; (7) Rate of replacements; (8) Average life = life expectancy of a new machine; (9) Index of replacement costs; (11) Limit of shape of possible other mortality curves. The opposite limit is (8).

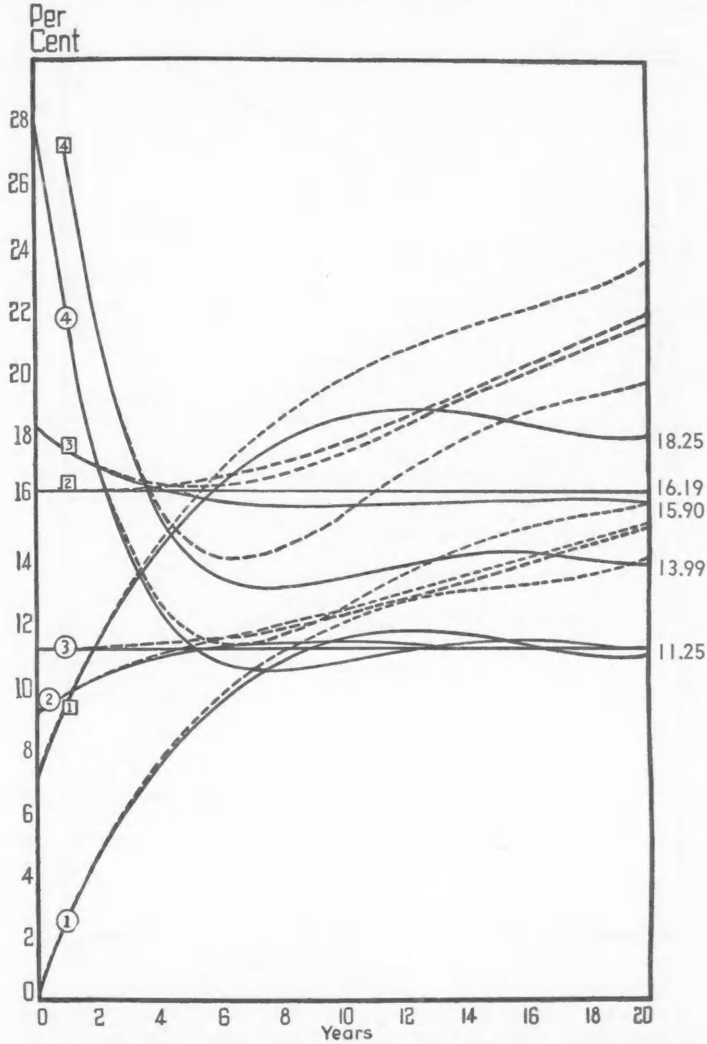

FIGURE 2.—Vertical scale represents per cent per annum of cost of original plant. Smooth lines indicate static conditions. Broken lines include increase in replacement costs at the force of 2 per cent per annum. Circles around numbers indicate rates of depreciation per annum. Squares around numbers indicate the sum of depreciation rates and a fair return at the force of 7 per cent per annum on the respective book values.

1. Retirement method; 2. Annuity method at the force of 7 per cent per annum; 3. Straight-line method; 4. Diminishing-balance method.

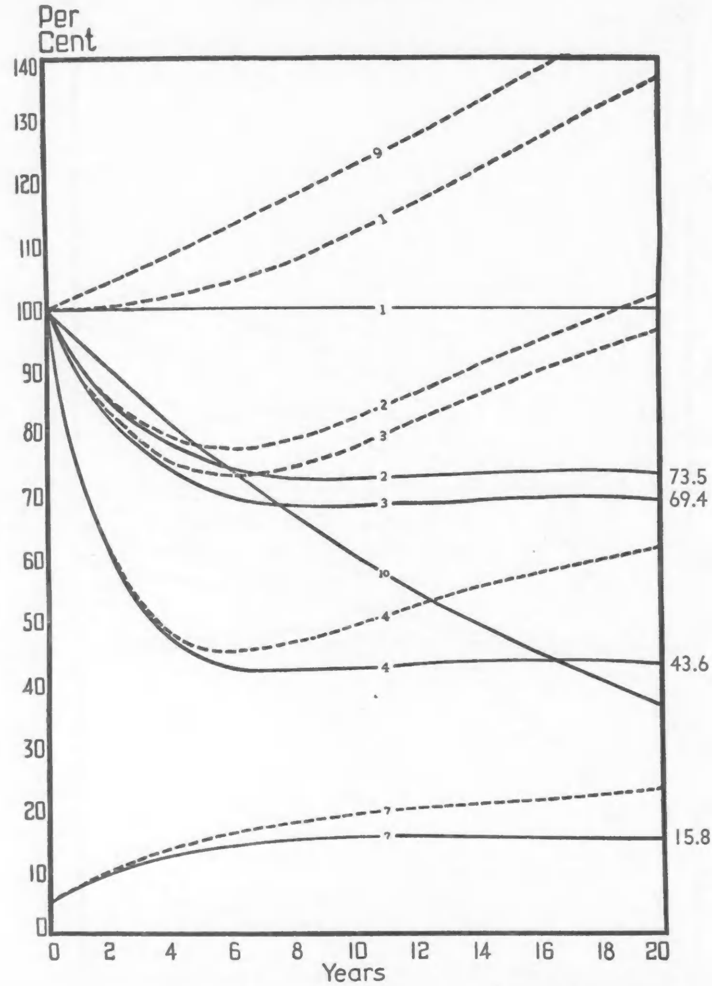

FIGURE 3.—Vertical scale represents per cent of cost of plant in service at any time *t*, when number of machines expands at the force of 5 per cent per annum and when replacement costs remain at the level of original costs. Smooth lines indicate static replacement costs and plant expansion at the force of 5 per cent per annum. Broken lines include increase in replacement costs at the force of 2 per cent per annum.

Composite-book-value curves: (1) Retirement method; (2) Annuity method at the force of 7 per cent per annum; (3) Straight-line method; (4) Diminishing-balance method.

Auxiliary curves: (7) Rate of replacements and additions; (9) Combined index of expansion and replacement costs; (10) Original cost of plant. True ordinates of all curves have been discounted at the force of 5 per cent per annum before plotting.

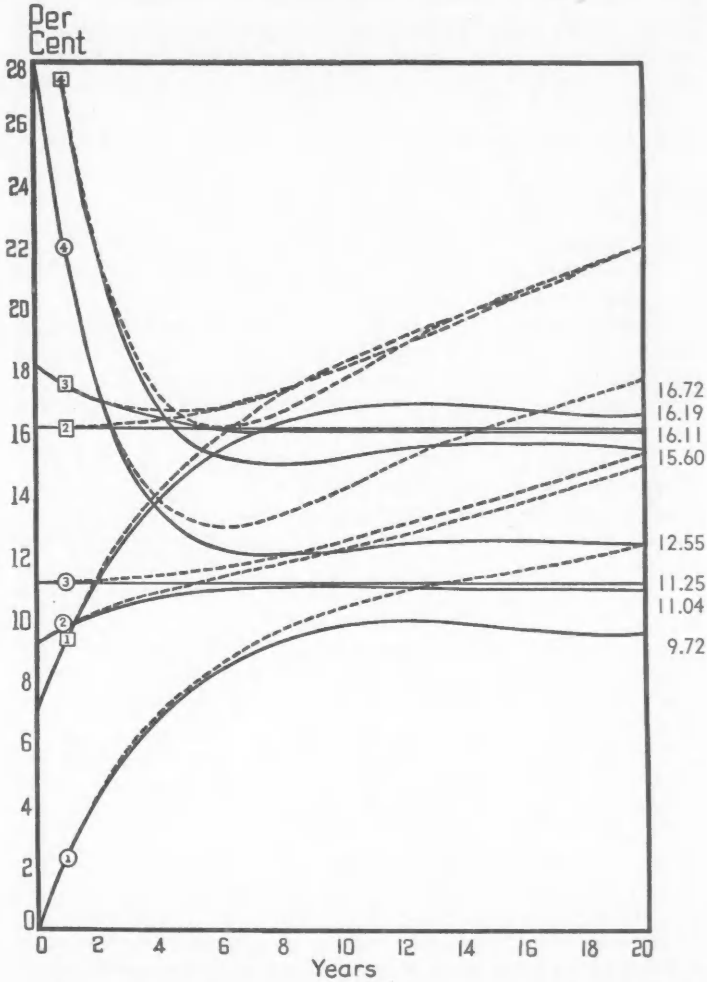

FIGURE 4.—Vertical scale represents per cent per annum of cost of original plant, chargeable at any time t per the original number of machines. Smooth lines indicate static replacement costs and plant expansion at the force of 5 per cent per annum. Broken lines include increase in replacement costs at the force of 2 per cent per annum. Circles around numbers indicate depreciation rates per annum. Squares around numbers indicate the sum of depreciation rates and a fair return at the force of 7 per cent per annum on the respective book values.

1. Retirement method; 2. Annuity method at the force of 7 per cent per annum; 3. Straight-line method; 4. Diminishing-balance method.

sented merely to outline the technique; book-value curves and depreciation rates can be compared far more effectively in graphic form. The four graphs submitted herewith illustrate the following cases:

A. Static conditions.
B. The number of machines in service remains constant. Replacement costs increase at the force of 2 per cent per annum.
C. The number of machines in service increases at the force of 5 per cent per annum. Replacement costs are static.
D. The number of machines in service increases at the force of 5 per cent per annum. Replacement costs increase at the force of 2 per cent per annum.

The scrap value is 10 per cent of the purchase price of a machine; the maximal age n is twenty years. The annuity method was calculated at the force of 7 per cent per annum. These assumptions still simplify actual conditions greatly, but they lead nevertheless to a number of important conclusions:

1. For any method, except the retirement method, the book value of an expanding plant stabilizes itself at a higher proportion of cost than that of a static plant. An expanding plant contains a higher proportion of relatively new machines; its average life expectancy is greater.

2. In the entirely static case, any method of depreciation will ultimately produce the same charge to operations, i.e., a charge at the rate of renewal b/a of the services consumed. The amount of profit reported by the books will ultimately be independent of the depreciation method. The same profit, however, will be measured per centum of widely differing bases.

3. When expansion is present, differences in depreciation charges will never disappear. Books kept by the retirement method will report the lowest charge and therefore the highest profit, while the diminishing-balance method leads to the opposite result. Simultaneously the highest profit will appear to be the lowest one, per centum of the investment reported, and vice versa.

4. When expansion proceeds at a constant rate, the difference between the amount of profit and the amount of expansion (reinvestment) reported by the books is ultimately the same, no matter what method of depreciation be employed.[8] This means that dividend policies ultimately become independent of the depreciation method. The same is true in point 2 above because, in the static case, the entire profit is available in cash.

5. In terms of the number of machines in service, the depreciation

[8] Formulae (34) to (41) show that $xC(x, \infty)e^{-x\infty} - D(x, \infty)e^{-x\infty} = b\chi + sx$ for any method.

rate of the straight-line method remains unaffected by expansion. In the case of the annuity method, the sum of depreciation and a fair return shows similar independence of both expansion and the age of the plant.

6. Until the renewal rate is stabilized, increasing replacement costs increase book values relatively faster for an expanding plant than for a static one. The proportionate difference between depreciated present reproduction cost (appraisal) and depreciated historical cost is smaller for an expanding plant than for a static one. This development occurs in addition to that described in point 1 above.

7. In terms of the number of machines in service, increasing replacement costs increase depreciation charges under any method, but discrepancies are also created in the same sense as by expansion.

8. Both expansion and increasing replacement costs speed the automatic stabilizing process, i.e., they decrease the relative amplitudes of oscillations from the start.

The graphs also present the case of regulated monopolies for the assumption that the force of 7 per cent per annum is a fair return on the "rate base." To illustrate this situation, curves have been plotted, which show the sum of depreciation and a fair return on the respective book values. The following additional conclusions may then be derived:

9. Under entirely static conditions, the "rates" which might be claimed as fair will differ widely according to the method of depreciation. The retirement method is the most lucrative in the long run.

10. Moderate expansion tends to decrease discrepancies in "rates" considered fair under various methods. Beyond a certain limit, the relative positions of "rate" levels will be reversed. For instance if, in the case C above, expansion were to exceed the force of 6 per cent per annum,[9] the diminishing-balance method would become the most lucrative one, while the retirement method would be at the bottom in that respect.

11. Increasing replacement costs also decrease discrepancies among various fair-"rate" concepts. Figure 4 shows that, in the circumstances assumed, a combination of 5 per cent expansion and 2 per cent increase in replacement costs ultimately calls for the same "rate" under any method of depreciation. If either rate were higher, the reversal of "rate" levels would also occur.

III

The foregoing comparison of depreciation methods shows how desirable it would be to find the "true" method. The search for truth

[9] Solve for x the equation

$$-D_r(x, \infty)e^{-x\infty}+i=-D_d(x, \infty)e^{-x\infty}+iC_d(x, \infty)e^{-x\infty}.$$

starts from the economic theory of capital value. According to fundamental principles, the capital value (use value) of a single machine consists of the discounted net rental, plus the discounted scrap value. Prof. Harold Hotelling[10] formulates the concept as follows:

$$(43) \qquad V(t) = \int_t^T [zQ(\tau) - E(\tau)]e^{-\int_t^\tau i(\nu)d\nu}d\tau + s(T)e^{-\int_t^T i(\nu)d\nu}.$$

In this notation, which has been changed somewhat to avoid conflict with that adopted elsewhere in this paper, $V(t)$ = unknown capital value of a single machine, z = unknown "theoretical selling price" of a unit of product, $Q(t)$ = known rate of output, $E(t)$ = known operating expenses,[11] T = unknown time at which the machine ought to be discarded, $s(t)$ = known scrap value, and $i(t)$ = known rate of interest.

In accordance with the main postulate that "everything in the owner's power will be done to make $V(t)$ a maximum, when $t>0$,"[12] T is determined by $dV(t)/dT=0$:

$$(44) \qquad zQ(T) - E(T) = i(T)s(T) - s'(T).$$

The machine must be scrapped, when the sales exceed the operating expenses by interest on the scrap value (less the rate of change of the scrap value, if any).

A third equation is needed to find the three unknowns $V(t)$, z, and T. Prof. Hotelling obtains it by rewriting (43) for $t=0$, "since," he says, "the value of a new machine $[V(O)]$ is its cost."[13]

This statement can no longer be accepted as generally valid. The connection between the cost and the capital value of a new machine is by no means a direct one. Its market price oscillates around a balancing point determined marginally by the least efficient pair of all producers and consumers of such machines. In the general case, therefore, capital value is apt to be greater than cost, although fluctuations may temporarily lead to the reverse. Be that as it may in a special case, it follows that, unless by pure accident the equality happens to be true, its assumption will always distort the problem. All three answers so obtained must necessarily be wrong.

The proper way to pass from the capital value $V(t)$ to the book value (unexpired cost) $B(t)$ of a single machine is to substitute the rate of

[10] "A General Mathematical Theory of Depreciation," *Journal of the American Statistical Association*, September, 1925.

[11] No definition of operating expenses is given, but the formula shows that, if the selling price is to be correct, the term must include every business expense except depreciation.

[12] *Op. cit.*, p. 343.

[13] *Ibid.*

profit $p(t)$ for the rate of interest. Although (43) remains the correct capital-value formula, we must use for depreciation purposes

$$B(t) = \int_t^T [zQ(\tau) - E(\tau)]e^{-\int_t^\tau p(\nu)d\nu}d\tau + s(T)e^{-\int_t^T p(\nu)d\nu}; \tag{45}$$

$$B(0) = \int_0^T [zQ(\tau) - E(\tau)]e^{-\int_0^\tau p(\nu)d\nu}d\tau + s(T)e^{-\int_0^T p(\nu)d\nu}; \tag{46}$$

$$zQ(T) - E(T) = p(T)s(T) - s'(T). \tag{47}$$

The machine will now be replaced as soon as it ceases to earn the rate of profit on its scrap value, unless the market price of scrap varies. It is true that the capital value of a single machine may thus be kept below the maximum, but it is more important to maximize the value of the whole chain of machines succeeding each other. Only when no replacement is intended, is it correct to use (44) in lieu of (47).[14]

The simplest depreciation problem is that of a regulated monopoly, where $p(t)$ is known as the legal rate of profit. This rate can be made the actual rate by establishing the principle that discrepancies between the two create accounts receivable or payable, to be liquidated in due course. The task is then fully defined and the three equations (45) to (47) can be solved without difficulty. For public utilities at least, it appears theoretically possible to prescribe a standard method of depreciation, which will enforce the spirit of regulation. Prof. Hotelling's numerical example[15] is a fair illustration of public-utility depreciation theory as applied to a single machine, if the legal rate of profit is used in lieu of his rate of interest. It would be even more instructive, had he not omitted the scrap value for the sake of brevity. The method, however, is not applicable to competitive conditions, nor does it take into account the theory of the composite plant.

When it is attempted to apply the results of single-machine analysis to many machines installed at the same time, it becomes apparent that, if the value theory of discarding is applicable,[16] all terms under the integral sign of the book-value formula (45) must be functions of the date of scrapping T, which is given by the reversed mortality formula $T = M^{-1}(y)$. Equation (45) may be so revised and then summed up with respect to the ordinate:

[14] A variable rate of profit may be considered to pass without discontinuity from one machine to the next. This means, however, that the scrapping date of the first machine depends upon all successive replacements.

[15] *Op. cit.*, p. 347.

[16] That need not always be the case. When a machine has an almost fixed total capacity for service and can render it with almost uniform, or even increasing efficiency, discarding will have a physical, rather than a value aspect. The date of discarding should then be known from experience.

$$(48) \quad c(t) = \int_0^{M(t)} \left[\int_t^T [z(T)Q(T,\tau) - E(T,\tau)] e^{p(t-\tau)} d\tau + s(T)e^{p(t-T)} \right] dy.$$

That would be the "true" book-value formula of the original plant, to be placed alongside equations (19) to (22) for substitution in the general expression (16) of the composite book value. The composite "rate" or selling price $Z(x, q, t)$ per unit of output can also be found by the usual method of deriving composite curves. Let

$$(49) \quad \xi(t) = \int_0^{M(t)} z(T)Q(T,t)dy \text{ and } \lambda(t) = \int_0^{M(t)} Q(T,t)dy,$$

$$(50) \quad Z(x, q, t) = \frac{\xi(t) + \int_0^t u(x, q, \tau)\xi(t-\tau)d\tau}{\lambda(t) + \int_0^t u(x, \tau)\lambda(t-\tau)d\tau}, \qquad t < n.$$

This solution is oversimplified in many respects, because a change in replacement costs presumably changes the mortality curve and therefore also the renewal function. Additional variations arise, when expenses are analyzed into their component parts, each of which behaves differently, some being related to the whole enterprise, rather than to individual machines. It is impossible to consider these complications within the scope of the present article.

IV

Let us now search for the "true" method of depreciation applicable to the broader field of competitive enterprise. There, the rate of profit $p(t)$ is unknown, so that further information is needed. The first impulse is to turn to the market for an additional equation. The market price of the machine's product at any given time is determined by the rate of demand corresponding to the rate of production of the industry, subject to the cushioning effect of inventories. Cost regulates production only in the long run by influencing replacement policies. In order to recover at least a portion of the purchase price, a machine once acquired must be operated, even if the owner finds himself on the wrong side of the theoretical margin. Although the market price of the product thus depends upon the rate of output of every enterprise in a certain field, it is virtually independent of the production costs of the machines already in service. This situation apparently prevails as long as the market price is in excess of the direct expenses which can be saved by stoppage. For purposes of a general theory, the market price z is therefore best considered a known function. How to find it, is not primarily a depreciation problem, except in special cases.[17]

[17] Unregulated monopoly is not a special case in this sense, because the market

Even if the market price of the product is known, however, the task is still indefinite, because any suitable combination of book value and rate of profit will satisfy the conditions imposed. The next step which suggests itself is to introduce the changing opportunities for bargains in the market for machines. Theory carried to extremes would then postulate perfect foresight of all economic conditions up to the death of the enterprise. On such a basis, the dates of replacement and the total production surpluses of successive machines could be calculated, but the distribution of the surpluses over the lives of the respective machines, i.e., the continuous time shape of the profit, remains undefined. No matter how far analysis and conjecture are carried, it is necessary to assume the form of the profit function either deliberately or by doing—perhaps unwittingly—something equivalent. *Any depreciation method ever devised amounts merely to such an assumption.*

To support this perhaps startling conclusion, Prof. Hotelling's "general" theory will be re-examined. Its essence is evidently:

$$\text{(51)} \qquad B(t) = \int_t^T [wQ(\tau) - E(\tau)] e^{-\int_t^T i(\nu) d\nu} d\tau + s(T) e^{-\int_t^T i(\nu) d\nu}.$$

At first sight, this may seem to be a general way to pass from capital value to book value, because the substitution of the so-called unit cost w for the unit selling price z apparently rectifies the error committed by forcing $V(t) = B(t)$. On closer scrutiny it is found, however, that if z is considered constant for illustrative purposes, it does not follow that w must also be constant. The assumption that it is amounts to a definition of the profit function in a manner which is inconsistent with the value theory of discarding.

The proof is easily furnished by equating the truly "general" expression (45) with the special assumption (51) and solving for the profit:

$$\text{(52)} \qquad p(t) = i(t) + \frac{(z - w)Q(t)}{B(t)} \cdot$$

When $z > w$, it follows immediately that $p(T) > i(T)$, because $B(T) > 0 < Q(T)$. The profit at the moment of discarding contradicts the theory upon which discarding is based!

price and the proportion of the maximal productive capacity which will yield the most remunerative market price can be found from two equations virtually independent of the other three. The first of these additional equations is evidently the demand function; the second expresses the requirement that the derivative of the net rental with respect to the unknown proportion of the maximal productive capacity must vanish.

The principle involved may be clarified further by citing a numerical example. Let $Q(t)=100e^{-.04t}$, $E(t)=\$10e^{.20t}$, $i=.05$, $B(0)=\$200$, and $B(T)=s(T)=\$10$. These data must be substituted in Prof. Hotelling's three equations, namely (51) as written, the same equation for $t=0$, and equation (44), in which w should be read for z. The solution is $T=7.4771$ years, $w=\$.608392$, and

$$(53) \qquad B(t) = \$675.99e^{-.04t} + \$66.67e^{.20t} - \$542.66e^{.05t}.$$

If it so happens that $z=w$, all is well, because $p(t)=i$. Such an accident can not be expected, however. Let us therefore assign another value to z, say \$.628699. Substitution into (52) leads to the result:

$$(54) \quad p(t) = .05 + \$2.0307/B(t)e^{.04t};\ p(0) = .0603 \text{ and } p(T) = .2006.$$

The books report that the machine is earning over 20 per cent per annum on its scrap value at the moment when it is discarded for the reason that it is earning only 5 per cent! All theories of depreciation known to me fail to pass this test. Yet it is not difficult to devise a method which will behave in a consistent manner when foresight is perfect. The simplest example would be:

$$(55) \quad B(t) = \$571.54e^{-.04t} + \$76.92e^{.20t} - \$448.46e^{.07t} \text{ for } 0 \leqq t \leqq 7.597$$

and

$$B(t) = \$10 \qquad \text{for } 7.597 \leqq t \leqq 7.6117.$$

This book-value curve determines the profit function

$$(56) \qquad p(t) = .07 \qquad \text{for } 0 \leqq t \leqq 7.597,$$

and

$$p(t) = 6.28699e^{-.04t} - e^{.20t} \qquad \text{for } 7.597 \leqq t \leqq 7.6117,$$

$$p(7.6117) = .05.$$

If this had been a public-utility problem with the added condition that the machine is not to be replaced, but allowed to serve until it ceases to earn the rate of interest on its scrap value, it could be said that equation (55) is the correct solution.[18] For the competitive case,[19] it looks reasonable, because it is at least consistent, but any number of other profit curves may be subjected to the same condition $p(7.6117) = .05$. Each will produce a different book-value curve, i.e., a different depreciation theory, none of which is either the "true" or the "general"

[18] The calculation consists of solving simultaneously the three equations (45), (46), and (47) for $p(t)=.07$. This leads to $T=7.597$, $z=\$.628699$ and to (55). The selling price was then substituted in equation (44) to find $T=7.6117$. The profit function for the interval $7.597 \leqq t \leqq 7.6117$ is given by equation (47).

[19] Consider z as known and solve instead for the unknown profit $p(t)=p$.

one. All that can be demanded is that the time shape of the profit and the date of discarding be based upon a single assumption, not upon two contradictory ones. If $p_j(t)$ be the rate of profit of the jth machine in the chain of replacements and T_j its date of replacement ($j=1, 2, 3, \cdots, \omega$), the condition $p_j(T_j)=p_{j+1}(T_j)$ must be observed until the end of operations, when it may be written $p_\omega(T_\omega)=i(T_\omega)$. Strictly speaking, $p_j'(T_j)=p_{j+1}'(T_j)$ is also desirable.

The purely theoretical value aspect of depreciation could, of course, be elaborated indefinitely. Foremost among the topics omitted are the question of obsolescence, i.e., gradual improvement in the type of machines used and the relationship between the rate of production and the productive capacity. All assumptions made are also subject to other more or less obvious limitations which cannot be enumerated here. As Walras has pointed out long ago, the solution of any economic problem ultimately resolves itself into the task of solving all other problems at the same time. In such circumstances, the last-comer is always in a strategic position. No matter how many simultaneous equations his predecessors may have set up, he can always prove their naïveté by pointing to another equation, which they ought not to have overlooked.

V

Outside the limited field of regulated monopoly, the search for the true depreciation method has apparently ended in failure, because the only clue which could be found is the rate of profit itself. This rate is necessarily indeterminate within the life periods of the machines succeeding each other. In any event, therefore, an arbitrary assumption is needed, the most highly developed form of which would be the fitting of a continuous curve to the successive blocks of production surpluses.

This situation is apt to turn the thoughts of the investigator into more practical channels. Theory is an insidious affliction, however, which is difficult to shake off. For instance, it sounds highly practical to say that no method of depreciation, haphazard, naïve, or absurd though it may be, can change the *true* value of a machine. To do that, something must be done to the machine, not to the books. The date of discarding affects the machine; depreciation is a matter of bookkeeping. Prof. Hotelling has made substantially this remark which is frequently echoed.[20]

The conclusion, unfortunately, is valid only in theory, when emphasis is placed on the word "true," which implies perfect foresight. As soon as this assumption contrary to fact is abandoned, the fallacy becomes apparent. It is best disclosed by a formula of capital value

[20] *Op. cit.*, p. 341.

based entirely on the books. By means of elementary operations, the capital-value formula (43) can easily be converted into

$$(57) \qquad V(t) = B(t) + \int_t^T [p(\tau) - i(\tau)]B(\tau)e^{-\int_t^\tau i(\nu)d\nu}d\tau.$$

Capital value equals the book value, plus the discounted excess profits.

Theoretically, this statement is true for any book value and any method of depreciation, whether based upon cost or not. In practice, it is the well-known formula for appraising the capital value of a business by past experience. The future date T, up to which the prospective purchaser is willing to look ahead, is determined largely by the rate of profit, on the ground that the more moderate it is, the longer it is likely to last. Bookkeeping accordingly has great influence upon capital value after all. The more "conservative" the book value, the lower the capital value is apt to be. That, of course, is no longer the theoretically true value, but it is the only one which has a practical significance.

Accountants immediately discard their own figures and demand an appraisal of the plant and other fixed assets, whenever they are called upon to compute capital value for the purposes of sale, reorganization, etc.[21] Apart from such occasions they adhere to their depreciation methods with the proviso that the method itself matters less than consistent adherence to it, once it has been adopted. These methods generally limit guessing to a minimum considered unavoidable in the circumstances. Guesses, so accountants feel, ought to be left to the stock market which will, in due time, accustom itself to the particular accounting methods followed by a given company.[22]

That, briefly, is the present status of depreciation theory and practice. The principal error, of which mathematical economists and accountants are equally guilty, is that both groups have so far examined the problem only in terms of a single machine, disregarding the actuarial theory of the composite plant, which introduces various new aspects. The first of these is mortality and replacement; attention to it is indispensable for the purpose of gaining a true perspective. Even more important, however, are the practical criteria of the capital value of a composite plant. In the absence of perfect foresight, appraisals by the stock market are based to a great extent upon a cursory analysis of the rate of profit in terms of its contributory causes, viz., opportunity, efficiency, risk, original book value (frequently not cost), and de-

[21] For a discussion of legal and accounting practices in determining capital value see Gabriel A. D. Preinreich, "The Law of Goodwill," *Accounting Review*, Dec., 1936 and "Goodwill in Accountancy," *Journal of Accountancy*, July, 1937.

[22] Cf. conclusions 2 and 4 reached at the end of Section II.

preciation method (frequently not all-inclusive). Apart from certain extraneous items, the value of the plant can not be distinguished from the value of an enterprise as a whole. It no longer depends upon the future services of the machines which compose the plant at any given moment, but also upon the services of those which will be installed in the future.[23] Depreciation theories must be formulated and judged in the light of the situation as it actually exists, not by assuming perfect foresight. What can be done to aid and simplify the appraisal function of the stock markets is the real problem, because market quotations are the most important manifestations of capital value.

New York, N. Y.

[23] This introduces the hitherto neglected concepts of the *expansion rate* and the *horizon* as important criteria of capital value. For a cursory discussion see Gabriel A. D. Preinreich, *The Nature of Dividends*, New York, 1935, p. 9 and appendix.

THE GENERAL WELFARE IN RELATION TO PROBLEMS OF TAXATION AND OF RAILWAY AND UTILITY RATES*

By HAROLD HOTELLING

IN THIS PAPER we shall bring down to date in revised form an argument due essentially to the engineer Jules Dupuit, to the effect that the optimum of the general welfare corresponds to the sale of everything at marginal cost. This means that toll bridges, which have recently been reintroduced around New York, are inefficient reversions; that all taxes on commodities, including sales taxes, are more objectionable than taxes on incomes, inheritances, and the site value of land; and that the latter taxes might well be applied to cover the fixed costs of electric power plants, waterworks, railroads, and other industries in which the fixed costs are large, so as to reduce to the level of marginal cost the prices charged for the services and products of these industries. The common assumption, so often accepted uncritically as a basis of arguments on important public questions, that "every tub must stand on its own bottom," and that therefore the products of every industry must be sold at prices so high as to cover not only marginal costs but also all the fixed costs, including interest on irrevocable and often hypothetical investments, will thus be seen to be inconsistent with the maximum of social efficiency. A method of measuring the loss of satisfactions resulting from the current scheme of pricing, a loss which appears to be extremely large, will emerge from the analysis. It will appear also that the inefficient plan of requiring that all costs, including fixed overhead, of an industry shall be paid out of the prices of its products is responsible for an important part of the instability which leads to cyclical fluctuations and unemployment of labor and other resources.

A railway rate is of essentially the same nature as a tax. Authorized and enforced by the government, it shares with taxes a considerable degree of arbitrariness. Rate differentials have, like protective tariffs and other taxes, been used for purposes other than to raise revenue. Indeed, the difference between rail freight rates between the same points, according as the commodity is or is not moving in international transport, has been used in effect to nullify the protective tariff. While it has not generally been perceived that the problems of taxation and those of railway rate making are closely connected, so that two independent bodies of economic literature have grown up, nevertheless the underlying unity is such that the considerations applicable to

* Presented at the meeting of the Econometric Society at Atlantic City, December 28, 1937, by the retiring president.

taxation are very nearly identical with those involved in proper rate making. This essential unity extends itself also to other rates, such as those charged by electric, gas, and water concerns, and to the prices of the products of all industries having large fixed costs independent of the volume of output.

I. THE CLASSICAL ARGUMENT

Dupuit's work of 1844 and the following years[1] laid the foundation for the use of the diagram of Figure 1 by Marshall and other econo-

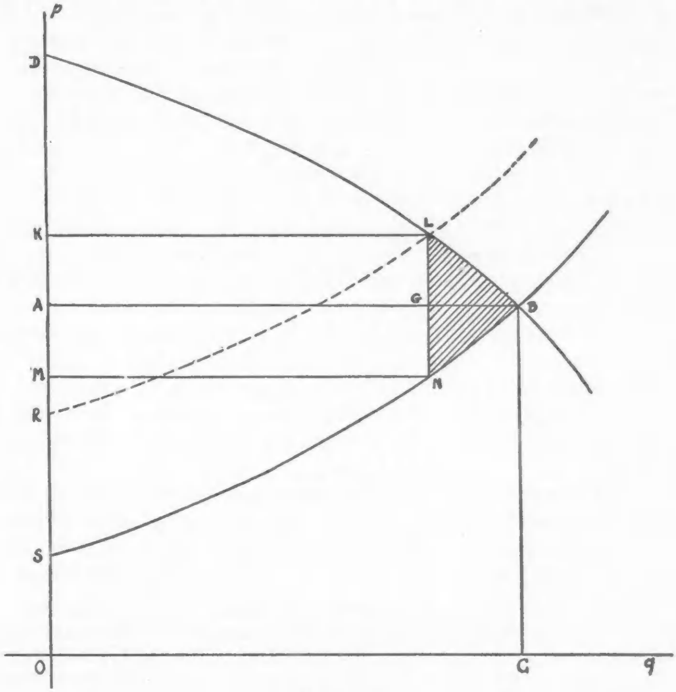

FIGURE 1

mists. A rising supply curve *SB* is used, and is sometimes regarded as coinciding with the marginal-cost curve. Such a coincidence would arise if there were free competition among producers, in the sense that each would regard the price as fixed beyond his control, and adjust

[1] Collected and reprinted with comments by Mario di Bernardi and Luigi Einaudi. "De l'Utilité et de sa Mesure," *La Riforma Soziale*, Turin, 1932.

his production so as to obtain maximum net profits. This condition is approximated, for example, in most agriculture. DB is a declining demand curve. The buyers are presumed to compete freely with each other. The actual quantity and price are the co-ordinates of the intersection B. Then it is supposed that a tax t per unit is imposed upon the sellers. Since this is a uniform increment to marginal cost, the marginal-cost curve SB is lifted bodily to the new position RL, at height $t=SR=NL$ above its former position.

Three conclusions have been derived with the help of this figure, all of which must be reviewed to take account of the interrelations of the particular commodity in question with others. One of these arguments has almost universally been accepted, but must be rejected when account is taken of related commodities. A second has been accepted, and is actually true. The third has been condemned and attacked by a long line of prominent economists, but in the light of the more thorough analysis made possible by modern mathematical methods must now in its essence be accepted. The first is the proposition that since the point L of intersection of the demand curve with the supply curve RL is higher by GL, a fraction of the tax rate NL, than the intersection B with the tax-free curve SB, therefore the price is increased as a result of the tax, by an amount less than the tax. That this conclusion is not necessarily true when account is taken of related commodities I have shown in an earlier paper.[2] The second proposition—whose conclusion remains valid under certain plausible assumptions[2a]—is that, since L is to the left of B, the quantity of the taxed commodity will diminish. With this diminution is associated an approximately measurable net social loss.

The third argument is based on Dupuit's, and is of primary concern here. Dupuit sought a criterion of the value to society of roads, canals, bridges, and waterworks. He pointed out the weakness of calling the value of a thing only what is paid for it, since many users would if

[2] "Edgeworth's Taxation Paradox and the Nature of Demand and Supply Functions," *Journal of Political Economy*, Vol. 40, 1932, pp. 577–616. Edgeworth had discovered, and maintained against the opposition of leading economists, that a monopolist controlling two products may after the imposition of a tax on one of them find it profitable to reduce both prices, besides paying the tax. However he regarded this as a "mere curiosum," unlikely in fact to occur, and peculiar to monopoly. But it is shown in the paper cited that the phenomenon is also possible with free competition, and is quite likely to occur in many cases, either under monopoly or under competition.

[2a] On p. 600 of the paper just cited the conclusion is reached that it is reasonable to regard the matrix of the quantities h_{ij} as negative definite. From this and equation (19) of that page it follows that a positive increment in the tax t_j on the jth commodity causes a negative increment in the quantity of this commodity.

necessary pay more than they actually do pay. The total benefit he measured by the aggregate of the maximum prices that would be paid for the individual small units of the commodity (a term used here to include services, e.g., of canals) corresponding to the costs of alternatives to the various uses. If $p=f(q)$ is the cost of the best alternative to the use of an additional small unit of the commodity when q units are already used, then, if q_0 units are used altogether,

$$(1) \qquad \int_0^{q_0} f(q)dq$$

is the total benefit, which Dupuit called *utilité*, resulting from the existence of the canal or other such facility making possible the commodity (service) in question. Since $p=f(q)$ is the ordinate of the demand curve DB in Figure 1, this total benefit is the total area under the arc DB. To obtain what is now called the *consumers' surplus* we must subtract the amount paid by consumers, namely the product of the price by the quantity, represented by the rectangle $OCBA$. Thus the consumers' surplus is represented by the curvilinear triangle ABD. There is also a *producers' surplus* represented by the lower curvilinear triangle SBA; this is the excess of the money received by producers (the area of the rectangle $OCBA$) over the aggregate of the marginal costs, which is represented by the curvilinear figure $OCBS$. The total net benefit, representing the value to society of the commodity, and therefore the maximum worth spending from the public funds to obtain it, is the sum of consumers' and producers' surpluses, and is represented by the large curvilinear triangle SBD. It is the difference between the integral (1) of the demand function and the integral between the same limits of the marginal-cost function.

Imposition of the tax, by raising the price to the level of KL, appears to reduce the consumers' surplus to the curvilinear area KLD. The new producers' surplus is the area RLK, which equals SNM. There is also a benefit on account of the government revenue, which is the product of the new quantity MN by the tax rate NL, and is therefore measured by the area of the rectangle $MNLK$. The sum of these three benefits is $SNLD$. It falls short of the original sum of producers' and consumers' surpluses by the shaded triangular area NBL.

This shaded area represents the net social loss due to the tax, and was discovered by Dupuit. If the tax is small enough, the arcs BL and NB may be treated as straight lines, and the area of the triangle is, to a sufficient approximation, half the product of the base NL by the altitude GB. Since GB is the decrement in the quantity produced and consumed because of the tax, and NL is the tax rate, we may say that the net loss resulting from the tax is half the product of the tax rate

by the decrement in quantity. But since the decrement in quantity is, for small taxes, proportional to the tax rate, it then follows that the net loss is proportional to the *square* of the tax rate. This fact also was remarked upon by Dupuit.

This remarkable conclusion has frequently been ignored in discussions in which it should, if correct, be the controlling consideration. The open attacks upon it seem all to be based on an excessive emphasis on the shortcomings of consumers' and producers' surpluses as measures of benefits. These objections are four in number: (1) Since the demand curve for a necessity might for very small quantities rise to infinity, the integral under the curve might also be infinite. This difficulty can be avoided by measuring from some selected value of q greater than zero. Since in the foregoing argument it is only *differences* in the values of the surpluses that are essentially involved, it is not necessary to assign exact values. The situation is the same as in the physical theory of the potential, which involves an arbitrary additive constant and so may be measured from any convenient point, since only its differences are important. (2) Pleasure is essentially nonmeasurable and so, it is said, cannot be represented by consumers' surplus or any other numerical magnitude. We shall meet this objection by establishing a generalized form of Dupuit's conclusion on the basis of a ranking only, without measurement, of satisfactions, in the way represented graphically by indifference curves. The same analysis will dispose also of the objections (3) that the consumers' surpluses arising from different commodities are not independent and cannot be added to each other, and (4) that the surpluses of different persons cannot be added.

In connection with the last two points, it will be observed that if we have a set of n related commodities whose demand functions are

$$p_i = f_i(q_1, q_2, \cdots, q_n), \qquad (i = 1, 2, \cdots, n),$$

then the natural generalization of the integral representing total benefit, of which consumers' surplus is a part, is the line integral

$$\int(f_1dq_1 + f_2dq_2 + \cdots + f_ndq_n), \tag{2}$$

taken from an arbitrary set of values of the q's to a set corresponding to the actual quantities consumed. The net benefit is obtained by subtracting from (2) a similar line integral in which the demand functions $f_1, f_2, \cdots, f_n$ are replaced by the marginal-cost functions

$$g_i(q_1, q_2, \cdots, q_n), \qquad (i = 1, 2, \cdots, n).$$

If we put

$$h_i = f_i - g_i,$$

the total net benefit is then measured by the line integral

$$w = \int \sum h_i dq_i. \tag{3}$$

Such indeterminacy as exists in this measure of benefit is only that which arises with variation of the value of the integral when the path of integration between the same end points is varied. The condition that all these paths of integration shall give the same value is that the integrability conditions

$$\frac{\partial h_i}{\partial q_j} = \frac{\partial h_j}{\partial q_i}$$

be satisfied. In the paper on "Edgeworth's Taxation Paradox" already referred to, and more explicitly in a later note,[3] I have shown that there is a good reason to expect these integrability conditions to be satisfied, at least to a close approximation, in an extensive class of cases. If they are satisfied, the surpluses arising from different commodities, and also the surpluses belonging to different persons, may be added to give a meaningful measure of social value. This breaks down if the variations under consideration are too large a part of the total economy of the person or the society in question; but for moderately small variations, with a stable price level and stable conditions associated with commodities not in the group, the line integral w seems to be a very satisfactory measure of benefits. It is invariant under changes in units of measure of the various commodities, and also under a more general type of change of our way of specifying the commodities, such as replacing "bread" and "beef" by two different kinds of "sandwiches." For these reasons the total of all values of w seems to be the best measure of welfare that can be obtained without considering the proportions in which the total of purchasing power is subdivided among individuals, or the general level of money incomes. The change in w that will result from a proposed new public enterprise, such as building a bridge, may fairly be set against the cost of the bridge to decide whether the enterprise should be undertaken. It is certainly a better criterion of social value than the aggregate $\sum p_i q_i$ of tolls that can be collected on various classes of traffic, as Dupuit pointed out for the case of a single commodity or service. The actual calculation of w in such a case would be a matter of estimation of vehicular and pedestrian traffic originating and terminating in particular zones, with a comparison of distances by alternative routes in each case, and an evaluation

[3] "Demand Functions with Limited Budgets," *Econometrica*, Vol. 3, 1935, pp. 66–78. A different proof is given by Henry Schultz in the *Journal of Political Economy*, Vol. 41, 1933, p. 478.

of the savings in each class of movement. Determination whether to build the bridge by calculation merely of the revenue $\sum p_i q_i$ obtainable from tolls is always too conservative a criterion. Such public works will frequently be of great social value even though there is no possible system of charging for their services that will meet the cost.

II. THE FUNDAMENTAL THEOREM

But without depending in any way on consumers' or producers' surpluses, even in the form of these line integrals, we shall establish a generalization of Dupuit's result. We take our stand on the firm ground of a system of preferences expressible by a function

$$\Phi = \Phi(q_1, q_2, \cdots, q_n)$$

of the quantities $q_1, q_2, \cdots, q_n$ of goods or services consumed by an individual per unit of time. If the function Φ, Pareto's *ophélimité*, has the same value for one set of q's as for another, then the one combination of quantities is as satisfactory to the individual in question as the other. For two commodities, Φ is constant along each of a set of "indifference curves"; and likewise for n commodities, we may think of a system of hypersurfaces of which one passes through each point of a space of n dimensions, whose Cartesian co-ordinates are the quantities of the various goods. These hypersurfaces we shall refer to as *indifference loci.*

It is to be emphasized that the indifference loci, unlike measures of pleasure, are objective and capable of empirical determination. One interesting experimental attack on this problem was made by L. L. Thurstone, who by means of questionnaires succeeded in mapping out in a tentative manner the indifference loci of a group of girls for hats, shoes, and coats.[4] Quite a different method, involving the study of actual family budgets, also appears promising.[5] The function Φ, on the other hand, is not completely determinable from observations alone, unless we are prepared to make some additional postulate about independence of commodities, as was done by Irving Fisher in defining utility,[6] and by Ragnar Frisch.[7] The present argument does not depend on any such assumption, and therefore allows the replacement of Φ by

[4] "The Indifference Function," *Journal of Social Psychology*, Vol. 2, 1931, pp. 139–167, esp. pp. 151 ff.

[5] R. G. D. Allen and A. L. Bowley, *Family Expenditure*, London, 1935.

[6] *Mathematical Investigations in the Theory of Value and Prices*, New Haven, 1892.

[7] *New Methods of Measuring Marginal Utility*, Tübingen, 1932. Dr. Frisch also considered the possibility of substitute commodities in his *Confluence Analysis*, and in collaboration with Dr. F. V. Waugh made an attempt to handle this situation statistically.

an arbitrary increasing function Ψ of Φ, such as sinh Φ, or $\Phi+\Phi^3$. The statements we shall make about Φ will apply equally to every such function Ψ. Negative values of the q's are the quantities of labor, or of goods or services, produced by the individual. It is with the understanding that this kind of indeterminacy exists that we shall sometimes refer to Φ and Ψ as utility functions.

Consider now a state in which income and inheritance taxes are used to pay for the construction of bridges, roads, railroads, waterworks, electric power plants, and like facilities, together with other fixed costs of industry; and in which the facilities may be used, or the products of industry consumed, by anyone upon payment of the additional net cost occasioned by the particular use or consumption involved in each case. This additional net cost, or marginal cost, will include the cost of the additional labor and other resources required for the particular item of service or product involved, beyond what would be required without the production of that particular item. Where facilities are not adequate to meet all demands, they are made so either by enlargement, or by checking the demand through inclusion in the price of a rental charge for the facilities, adjusted so as to equate demand to supply. Such a rental cost, of which the site rental of land is an example, is an additional source of revenue to the state; it must not be confused with carrying charges on invested capital, or with overhead cost. Some such charge is necessary to discriminate economically among would-be users of the facilities. Another example is that of water in a dry country; if demand exceeds supply, and no enlargement of supply is possible, a charge must be made for the water sufficient to reduce the demand to the supply. Such a charge is an element of marginal cost as here defined.

The individual retains, after payment of taxes, a money income m. At prices $p_1, p_2, \cdots, p_n$ determined in the foregoing manner, he can buy or sell such quantities $q_1, q_2, \cdots, q_n$ as he pleases, subject to the condition that

$$(4) \qquad \sum p_i q_i = m.$$

The combination he chooses will be such as to make his indifference function Φ a maximum, subject to the condition (4). We may put aside as infinitely improbable—having probability zero, though not impossible—the contingency that two different sets of values of the q's satisfying (4) will give the same degree of satisfaction. We therefore have that, if $q_1, \cdots, q_n$ are the quantities chosen under these conditions, and if $q_1', \cdots, q_n'$ are any other set of quantities satisfying (4), so that

$$(5) \qquad \sum p_i q_i' = m,$$

then

$$\Phi = \Phi(q_1, \cdots, q_n) > \Phi(q_1', \cdots, q_n') = \Phi + \delta\Phi,$$

say. Hence, putting $q_i' = q_i + \delta q_i$ in (5) and subtracting (4), we find that any set of values of $\delta q_1, \cdots, \delta q_n$ satisfying

$$\sum p_i \delta q_i = 0, \tag{6}$$

and not all zero, must have the property that

$$\delta\Phi = \Phi(q_1 + \delta q_1, \cdots, q_n + \delta q_n) - \Phi(q_1, \cdots, q_n) < 0. \tag{7}$$

Let us now consider an alteration of the system by the imposition of excise taxes and reduction of income taxes. Some of the taxes may be negative; that is, they may be bounties or subsidies to particular industries; or, instead of being called taxes, they may be called tolls, or charges for services or the use of facilities over and above marginal cost. There ensues a redistribution of production and consumption. Let p_i, q_i, and m be replaced respectively by

$$p_i' = p_i + \delta p_i, \quad q_i' = q_i + \delta q_i, \quad m' = m + \delta m, \tag{8}$$

where the various increments δp_i, δq_i are not constrained to be either positive or negative; some may have one sign and some the other. The yield of the new excise taxes will be the sum, over all individuals, of the quantity which for the particular individual we are considering is $\sum q_i' \delta p_i$. (We use the sign $\sum$ to denote summation over all commodities, including services.) Since this person's income tax is reduced by δm, the net increment of government revenue

$$\delta r = \sum q_i' \delta p_i - \delta m \tag{9}$$

may be imputed to him, in the sense that summation of δr over all persons gives the total increment of government revenue.[8] We neglect changes in administrative costs and the like.

The individual's budgetary limitation now takes the form $\sum p_i' q_i' = m'$, which may also be written

$$\sum (p_i + \delta p_i)(q_i + \delta q_i) = m + \delta m. \tag{10}$$

[8] A friendly critic writes. "It is not clear to me why δp_i should be the exact per-unit revenue of the state from an excise tax which raises the price by δp_i from its old level. . . . I should expect (referring to Figure 1) an increase in price of GL, and a revenue to the state of NL." The answer to this is that the summation of δr over all persons includes the sellers as well as the buyers, and that the government revenue per unit of the commodity is derived in part from each—though it must be understood that the contribution of either or both may be negative. In the classical case represented by Figure 1, the buyers' δp is the height GL, while the sellers' is NG in magnitude and is negative. Since q' is positive for the buyer and negative for the seller, the product $q'\delta p$ is in each case positive. The aggregate of these positive terms is the total tax revenue from the commodity.

Subtracting the budget equation (4) corresponding to the former system and using (8) we find that

$$(11) \qquad \delta m = \sum q_i'\delta p_i + \sum p_i\delta q_i.$$

Substituting this in (9) we find that

$$(12) \qquad \delta r = - \sum p_i\delta q_i.$$

Suppose that, to avoid disturbing the existing distribution of wealth, the excise taxes paid by each individual (in the sense of incidence just defined; not in the sense of handing over the money to the government in person) are exactly offset by the decrement in his income tax. Then $\delta r = 0$. From (12) it then follows that (6) is satisfied. Except in the highly improbable case of all the δq's coming out exactly zero, it would then follow from (7) that this man's new state is worse than his old. The change from income to excise taxes has resulted in a net loss of satisfactions. Conversely, if we start from a system of excise taxes, or any system in which sales are not at marginal cost, this argument shows that there is a possible distribution of personal income taxes such that everyone will be better satisfied to change to the system of income taxes with sales at marginal cost. The problem of the distribution of wealth and income among persons or classes is not involved in this proposition.

This argument may be expressed in geometrical language as follows: Let $q_1, \cdots, q_n$ be Cartesian coordinates in a space of n dimensions. Through each point of this space passes a hypersurface whose equation may be written $\Phi(q_1, \cdots, q_n) = \text{constant}$. The individual's satisfaction is enhanced by moving from one to another of these hypersurfaces if the value of the constant on the right side of the equation is thereby increased; this will usually correspond to moving in a direction along which some or all of the q's increase. The point representing the individual's combination of goods is however constrained in the first instance to lie in the hyperplane whose equation is (4). In this equation the p's and m are to be regarded as constant coefficients, while the q's vary over the hyperplane. A certain point Q on this hyperplane will be selected, corresponding to the maximum taken by the function Φ, subject to the limitation (4). If the functions involved are analytic, Q will be the point of tangency of the hyperplane with one of the "indifference loci." The change in the tax system means that the individual must find a point Q' in the new hyperplane whose equation is $\sum p_i'q_i = m'$. If we denote the coordinates of Q' by $q_1', \cdots, q_n'$, we have, upon substituting them in the equation of this new hyperplane, $\sum p_i'q_i' = m'$. If the changes in prices and m are such as to leave the government revenue unchanged, (12) must vanish; that is,

$$\sum p_i q_i' = \sum p_i q_i.$$

Since $\sum p_i q_i = m$, this shows that $\sum p_i q_i' = m$; that is, that Q' lies on the same hyperplane to which Q was confined in the first place. But since Q was chosen among all the points on this hyperplane as the one lying on the outermost possible indifference locus, for which Φ is a maximum, and since we are putting aside the infinitely improbable case of there being other points on the hyperplane having this maximizing property, it follows that Q' must lie on some other indifference locus, and that this will correspond to a lesser degree of satisfaction.

The fundamental theorem thus established is that *if a person must pay a certain sum of money in taxes, his satisfaction will be greater if the levy is made directly on him as a fixed amount than if it is made through a system of excise taxes which he can to some extent avoid by rearranging his production and consumption.* In the latter case, the excise taxes must be at rates sufficiently high to yield the required revenue *after* the person's rearrangement of his budget. The redistribution of his production and consumption then represents a loss to him without any corresponding gain to the treasury. This conclusion is not new. What we have done is to establish it in a rigorous manner free from the fallacious methods of reasoning about one commodity at a time which have led to false conclusions in other associated discussions.

The conclusion that a fixed levy such as an income or land tax is better for an individual than a system of excise taxes may be extended to the whole aggregate of individuals. In making this extension it is necessary to neglect certain interactions among the individuals that may be called "social" in character, and are separate and distinct from the interactions through the economic mechanisms of price and exchange. An example of such "social" interactions is the case of the drunkard who, after adjusting his consumption of whisky to what he considers his own maximum of satisfaction, beats his wife, and makes his automobile a public menace on the highway. The restrictive taxation and regulation of alcoholic liquors and certain other commodities do not fall under the purview of our theorems because of these social interactions which are not economic in the strict sense. With this qualification, and neglecting also certain possibilities whose total probability is zero, we have:

If government revenue is produced by any system of excise taxes, there exists a possible distribution of personal levies among the individuals of the community such that the abolition of the excise taxes and their replacement by these levies will yield the same revenue while leaving each person in a state more satisfactory to himself than before.

It is in the sense of this theorem that we shall in later sections

speak of "the maximum of total satisfactions" or "the maximum of general welfare" or "the maximum national dividend" requiring as a necessary, though not sufficient, condition that the sale of goods shall be without additions to price in the nature of excise taxes. These looser expressions are in common use, and are convenient; when used in this paper, they refer to the proposition above, which depends only on rank ordering of satisfactions; there is no connotation of adding utility functions of different persons.

The inefficiency of an economic system in which there are excise taxes or bounties, or in which overhead or other charges are paid by excesses of price over marginal cost, admits of an approximate measure when the deviations from the optimum system described above are not great, if, as is customary in this and other kinds of applied mathematics, we assume continuity of the indifference function and its derivatives. Putting for brevity

$$\Phi_i = \frac{\partial \Phi}{\partial q_i}, \qquad \Phi_{ij} = \frac{\partial^2 \Phi}{\partial q_i \partial q_j},$$

we observe that the maximum of Φ, subject to the budget equation (4), requires that

$$\Phi_i = \lambda p_i, \qquad (i = 1, 2, \cdots, n), \tag{13}$$

where the Lagrange multiplier λ is the marginal utility of money. Differentiating this equation gives

$$\Phi_{ij} = \lambda \frac{\partial p_i}{\partial q_j} + p_i \frac{\partial \lambda}{\partial q_j} \cdot \tag{14}$$

Expanding the change in the utility or indifference function we obtain, with the help of (13), (12), and (14),

$$\begin{aligned} \delta\Phi &= \sum \Phi_i \delta q_i + \tfrac{1}{2} \sum \sum \Phi_{ij} \delta q_i \delta q_j + \cdots \\ &= -\lambda \delta r + \tfrac{1}{2}\lambda \sum\sum \frac{\partial p_i}{\partial q_j} \delta q_i \delta q_j - \tfrac{1}{2} \delta r \sum \frac{\partial \lambda}{\partial q_j} \delta q_j + \cdots \end{aligned} \tag{15}$$

where the terms omitted are of third and higher order, and are therefore on our assumptions negligible. Their omission corresponds to Dupuit's deliberate neglect of curvilinearity of the sides of the shaded triangle in Figure 1. With accuracy of this order we have further,

$$\delta p_i = \sum_j \frac{\partial p_i}{\partial q_j} \delta q_j, \qquad \delta \lambda = \sum_j \frac{\partial \lambda}{\partial q_j} \delta q_j.$$

Upon substituting for these expressions, (15) reduces to

$$\delta\Phi = -\lambda \delta r + \tfrac{1}{2} \lambda \sum \delta p_i \delta q_i - \tfrac{1}{2} \delta r \delta \lambda + \cdots. \tag{16}$$

If the readjustment from the original state of selling only at marginal cost, with income taxes to pay overhead, is such as to leave $\delta r=0$ as above, (16) reduces to

$$\delta\Phi = \tfrac{1}{2}\lambda\sum \delta p_i \delta q_i + \cdots, \tag{17}$$

where the terms omitted are of higher order.

As another possibility we may consider a substitution of excise for income tax so arranged as to leave this person's degree of satisfaction unchanged. Upon putting $\delta\Phi=0$ in (16) and solving for δr we have, apart from terms of higher order,

$$\delta r = \tfrac{1}{2}\sum \delta p_i \delta q_i + \cdots. \tag{18}$$

This is the net loss to the state in terms of money, so far as this one individual is concerned. The net loss in terms of satisfactions is merely the product of (18) by the marginal utility of money λ, that is, (17), if we neglect terms of higher order than those written. The total net loss of state revenue resulting from abandonment of the system of charging only marginal costs, and uncompensated by any gain to any individual, is the sum of (18) over all individuals. If the prices are the same for all, this sum is of exactly the same form as the right-hand member of (18), with δq_i now denoting the increment (positive or negative) of the total quantity of the ith commodity.

The approximate net loss

$$\tfrac{1}{2}\sum \delta p_i \delta q_i \tag{19}$$

may be regarded as the sum of the areas of the shaded triangles in the older graphic demonstration. It should however be remembered that the readjustment of prices caused by excise taxes is not necessarily in the direction formerly supposed, that some of the quantities and some of the prices may increase and some decrease, and that some of the terms of the foregoing sum may be positive and some negative. But the aggregate of all these varying terms is seen by the foregoing argument to represent a dead loss, and never a gain, as a result of a change from income to excise taxes, or away from a system of sales at marginal cost. Any inaccuracy of the measure (19) is of only the same order as the error involved in replacing the short arcs LB and NB in Figure 1 by straight segments, and can never affect the sign.

It is remarkable, and may appear paradoxical, that without assuming any particular measure of utility or any means of comparison of one person's utility with another's, we have been able to arrive at (19) as a valid approximation measuring in money a total loss of satisfactions to many persons. That the result depends only on the conception of ranking, without measurement, of satisfactions by each person is readily apparent from the foregoing demonstration; or we may for any

person replace Φ by another function Ψ as an index of the same system of ranks among satisfactions. If we do this in such a way that the derivatives are continuous, we shall have $\Psi=F(\Phi)$, where F is an increasing function with continuous derivatives. Upon writing the expressions for the first and second derivatives of Ψ in terms of those of F and Φ it may be seen that the foregoing formulae involving Φ are necessary and sufficient conditions for the truth of the same equations with Ψ written in place of Φ. The result (18) is independent of which system of indicating ranks is used. The fundamental fact here is that *arbitrary* analytic transformations, even of very complicated functional forms, always induce *homogeneous linear* transformations of differentials.

Not only the approximation (19) but also the whole expression indicated by (18) are absolutely invariant under all analytic transformations of the utility functions of all the persons involved. These expressions depend only on the demand and supply functions, which are capable of operational determination. They represent simply the money cost to the state of the inefficiency of the system of excise taxation, when this is arranged in such a way as to leave unchanged the satisfactions derived from his private income by each person.

The arguments based on Figure 1 have been repeated with various degrees of hesitation, or rediscovered independently, by numerous writers including Jevons, Fisher, Colson, Marshall, and Taussig. Marshall considered variations of the figure involving downward-sloping cost curves and multiple solutions, and was led to the proposal (less definite than that embodied in the criterion established by our theorem) that incomes and increasing-cost industries be taxed to subsidize decreasing-cost ones. He observed the difficulty of defining demand curves and consumers' surplus in view of the interdependence of demand for various commodities. These difficulties are indeed such that it now seems better to stop talking about demand *curves*, and to substitute demand *functions*, which will in general involve many variables, and are not susceptible of graphic representation in two or three dimensions. Marshall was one of those misled by Figure 1 into thinking that a tax of so much per unit imposed on producers of a commodity leads necessarily to an increase of price by something less than the tax.

Though the marginal-cost curve in Figure 1 slopes upward, no such assumption is involved in the present argument. It is perfectly possible that an industry may be operated by the state under conditions of diminishing marginal cost. The criterion for a small increase in production is still that its cost shall not exceed what buyers are willing to pay for it; that is, the general welfare is promoted by offering it for

sale at its marginal cost. It may be that demand will grow as prices decline until marginal cost is pushed to a very low level, far below the average cost of all the units produced. In such a case the higher cost of the first units produced is of the same character as fixed costs, and is best carried by the public treasury without attempting to assess it against the users of the particular commodity as such. Our argument likewise makes no exception of cases in which more than one equilibrium is possible. Where there are multiple solutions we have that sales at marginal cost are a necessary, though not a sufficient, condition for the optimum of general welfare.

The confusion between marginal and average cost must be avoided. This confusion enters into many of the arguments for laissez-faire policies. It is frequently associated with the calm assumption, as a self-evident axiom, that the whole costs of every enterprise must be paid out of the prices of its products. This fallacious assumption appears, for example, in recent writings on government ownership of railroads. It has become so ingrained by endless repetition that it is not even stated in connection with many of the arguments it underlies.

III. TAX SYSTEMS MINIMIZING DEAD LOSS

The magnitude of the dead loss varies greatly according to the objects taxed. While graphic arguments are of suggestive value only, it may be observed from Figure 1 that the ratio of the dead loss NBL to the revenue $MNLK$ depends greatly on the slopes of the demand and supply curves in the neighborhood of the equilibrium point B. It appears that if either the demand or the supply curve is very steep in this neighborhood, the dead loss will be slight. For a tax on the site rental value of land, whose supply curve is vertical, the dead loss drops to zero. A tax on site values is therefore one of the very best of all possible taxes from the standpoint of the maximum of the total national dividend. It is not difficult to substantiate this argument in dealing with related commodities; for the δq_i's corresponding to such a tax are zero. Since the incidence is on the owner of the land and cannot be shifted by any readjustment of production, it has the same advantages as an income tax from the standpoint of maximizing the national dividend. The fact that such a land tax cannot be shifted seems to account for the bitterness of the opposition to it. The proposition that there is no ethical objection to the confiscation of the site value of land by taxation, if and when the nonlandowning classes can get the power to do so, has been ably defended by H. G. Brown.[9]

Land is the most obviously important, but not by any means the only good, whose quantity is nearly or quite unresponsive to changes

[9] *The Theory of Earned and Unearned Incomes*, Columbia, Missouri, 1918.

in price, and which is not available in such quantities as to satisfy all demands. Holiday travel sometimes leads to such a demand for the use of railroad cars as to bring about excessive and uncomfortable crowding. If the total demand the year around is not sufficiently great to lead to the construction of enough more cars to relieve the crowding, the limited space in the existing cars acquires a rental value similar to that of land. Instead of selling tickets to the first in a queue, or selling so many as to bring about an excessive crowding that would neutralize the pleasure of the holiday, the economic way to handle this situation would be to charge a sufficiently high price to limit the demand. The revenue thus obtained, like the site value of land, may properly be taken by the state. The fact that it helps to fill the treasury from which funds are drawn to pay for replacement of the cars when they wear out, and to cover interest on their cost in the meantime, does not at all mean that any attempt should be made to equate the revenue from car-space rental to the cost of having the cars in existence.

Another thing of limited quantity for which the demand exceeds the supply is the attention of people. Attention is desired for a variety of commercial, political, and other purposes, and is obtained with the help of billboards, newspaper, radio, and other advertising. Expropriation of the attention of the general public and its commercial sale and exploitation constitute a lucrative business. From some aspects this business appears to be of a similar character to that of the medieval robber barons, and therefore to be an appropriate subject for prohibition by a state democratically controlled by those from whom their attention is stolen. But attention attracting of some kinds and in some degree is bound to persist; and where it does, it may appropriately be taxed as a utilization of a limited resource. Taxation of advertising on this basis would be in addition to any taxation imposed for the purpose of diminishing its quantity with a view to restoring the property of attention to its rightful owners.

If for some reason of political expediency or civil disorders it is impossible to raise sufficient revenue by income and inheritance taxes, taxes on site values, and similar taxes which do not entail a dead loss of the kind just demonstrated, excise taxes may have to be resorted to. The problem then arises of so arranging the rates on the various commodities as to raise the required sum while making the total dead loss a minimum. A solution of this theoretical question, taking account of the interrelations among commodities, is given on p. 607 of the study of Edgeworth's taxation paradox previously referred to.

IV. EFFECT ON DISTRIBUTION OF WEALTH

We have seen that, if society should put into effect a system of sales at marginal cost, with overhead paid out of taxes on incomes, in-

heritances, and the site value of land, there would exist a possible system of compensations and collections such that everyone would be better off than before. As a practical matter, however, it can be argued in particular cases that such adjustments would not in fact be made; that the general well-being would be purchased at the expense of sacrifices by some; and that it is unjust that some should gain at the expense of others, even when the gain is great and the cost small. For example, it appears that the United States Government can by introducing cheap hydroelectric power into the Tennessee Valley raise the whole level of economic existence, and so of culture and intelligence, in that region, and that the benefits enjoyed by the local population will be such as to exceed greatly in money value the cost of the development, taking account of interest. But if the government demands for the electricity generated a price sufficiently high to repay the investment, or even the interest on it, the benefits will be reduced to an extent far exceeding the revenue thus obtained by the government. It is even possible that no system of rates can be found that will pay the interest on the investment; yet the benefits may at the same time greatly exceed this interest in value. It appears to be good public policy to make the investment, and to sell the electric energy at marginal cost, which is extremely small. But this will mean that the cost will have to be paid in part by residents of other parts of the country, in the form of higher income and inheritance taxes. Those who are insistent on avoiding a change in the distribution of wealth at all costs will object.

One answer to this objection is that the benefits from such a development are not by any means confined to the persons and the region most immediately affected. Cheap power leads for example to production of cheap nitrates, which cut down the farmers' costs even in distant regions, and may benefit city dwellers in other distant regions. A host of other industries brought into being by cheap hydroelectric power have similar effects in diffusing general well-being. There is also the benefit to persons who on account of the new industrial development find that they can better themselves by moving into the Tennessee Valley, or by investing their funds there. Furthermore, the nation at large has a stake in eradicating poverty, with its accompaniments of contagious diseases, crime, and political corruption, wherever these may occur.

A further answer to the objection that benefits may be paid for by those who do not receive them when such a development as that of the Tennessee Valley is undertaken is that no such enterprise stands alone. A government willing to undertake such an enterprise is, for the same reasons, ready to build other dams in other and widely scattered places, and to construct a great variety of public works. Each of these entails

benefits which are diffused widely among all classes. A rough randomness in distribution should be ample to ensure such a distribution of benefits that most persons in every part of the country would be better off by reason of the program as a whole.

If new electric-power, railroad, highway, bridge, and other developments are widely undertaken at public expense, always on the basis of the criterion of maximizing total benefits, the geographical distribution of the benefits, and also the distribution among different occupational, racial, age, and sex groups, would seem pretty clearly to be such that every such large group would on the whole be benefited by the program. There are, however, two groups that might with some reason expect not to benefit. One of these consists of the very wealthy. Income and inheritance taxes are likely to be graduated in such a way that increases in government spending will be paid for, both directly and ultimately, by those possessed of great wealth, more than in the proportion that the number of such persons bears to the whole population. It would not be surprising if the benefits received by such persons as a result of the program of maximum total benefit should fall short of the cost to them.

The other class that might expect not to benefit from such a program consists of land speculators. If we consider for example a bridge, it is evident that the public as a whole must pay a certain cost of construction, whether the bridge be paid for by tolls or by taxes on the site value of land in the vicinity. There will be much more use of the bridge if there are no tolls, so that the public as a whole will get more for its money if it pays in the form of land taxes. But it will not in general be possible to devise a system of land taxes that will leave everyone, without exception, in a position as good as or better than as if the bridge had not been built and the taxes had not been levied. Landowners argue that the benefits of the bridge go to others, not to them; and even in cases in which land values have been heightened materially as a result of a new bridge, the landowners have been known to be vociferous in favor of a toll system. Payment for the bridge by tolls (when this is possible) has the advantage that no one seems to be injured, since each one who pays to cross the bridge has the option of not using it, and is in that case as well off as if the bridge did not exist. This reasoning is not strictly sound, since the bridge may have put out of business a ferry which for some users was more convenient and economical. Nevertheless, it retains enough cogency to stiffen the resistance of real-estate interests to the more economical system of paying for the bridge by land taxes.

Attempts at excessive accuracy in assessing costs of public enterprises according to benefits received tend strongly to reduce the total

of those benefits, as in the case of the bridge. The welfare of all is promoted rather by a generous support of projects for communal spending in ways beneficial to the public at large, without attempting to recover from each enterprise its cost by charges for services rendered by that enterprise. The notion that public projects should be "self-liquidating," on which President Hoover based his inadequate program for combating the oncoming depression, while attractive to the wealthier taxpayers, is not consistent with the nation's getting the maximum of satisfactions for its expenditure.

V. DISTINCTION OF OPTIMUM FROM COMPETITIVE CONDITIONS

The idea that all will be for the best if only competition exists is a heritage from the economic theory of Adam Smith, built up at a time when agriculture was still the dominant economic activity. The typical agricultural situation is one of rising marginal costs. Free competition, of the type that has usually existed in agriculture, leads to sales at marginal cost, if we now abstract the effects of weather and other uncertainty, which are irrelevant to our problem. Since we have seen that sales at marginal cost are a condition of maximum general welfare, this situation is a satisfactory one so far as it goes. But the free competition associated with agriculture, or with unorganized labor, is not characteristic of enterprises such as railroads, electric-power plants, bridges, and heavy industry. It is true that a toll bridge may be in competition with other bridges and ferries; but it is a very different kind of competition, more in the nature of duopoly. To rely on such competition for the efficient conduct of an economic system is to use a theorem without observing that its premises do not apply. Free competition among toll-bridge owners, of the kind necessary to make the conclusion applicable, would require that each bridge be parallelled by an infinite number of others immediately adjacent to it, all the owners being permanently engaged in cutthroat competition. If the marginal cost of letting a vehicle go over a bridge is neglected, it is clear that under such conditions the tolls would quickly drop to zero and the owners would retire in disgust to allow anyone who pleased to cross free.

The efficient way to operate a bridge—and the same applies to a railroad or a factory, if we neglect the small cost of an additional unit of product or of transportation—is to make it free to the public, so long at least as the use of it does not increase to a state of overcrowding. A free bridge costs no more to construct than a toll bridge, and costs less to operate; but society, which must pay the cost in some way or other, gets far more benefit from the bridge if it is free, since in this case it will be more used. Charging a toll, however small, causes some

people to waste time and money in going around by longer but cheaper ways, and prevents others from crossing. The higher the toll, the greater is the damage done in this way; to a first approximation, for small tolls, the damage is proportional to the square of the toll rate, as Dupuit showed. There is no such damage if the bridge is paid for by income, inheritance, and land taxes, or for example by a tax on the

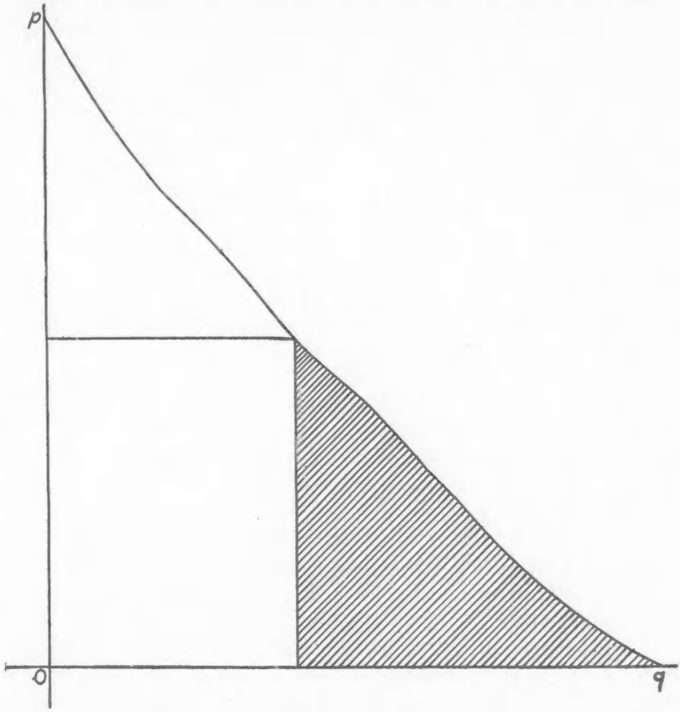

FIGURE 2

real estate benefited, with exemption of new improvements from taxation, so as not to interfere with the use of the land. The *distribution* of wealth among members of the community is affected by the mode of payment adopted for the bridge, but not the total wealth, except that it is diminished by bridge tolls and other similar forms of excise. This is such plain common sense that toll bridges have now largely disappeared from civilized communities. But New York City's bridge and tunnels across the Hudson are still operated on a toll basis, be-

cause of the pressure of real estate interests anxious to shift the tax burden to wayfarers, and the possibility of collecting considerable sums from persons who do not vote in the city.

If we ignore the interrelations of the services of a bridge with other goods, and also the slight wear and tear on the bridge due to its use, we may with Dupuit represent the demand for these services by a curve such as that in Figure 2. The total benefit from the bridge is then represented by the whole area enclosed between the demand curve and the axes, provided the bridge is free. All this benefit goes to users of the bridge. But if a toll is charged, of magnitude corresponding to the height of the horizontal line, the recipients of the toll are benefited to an extent represented by the area of the rectangle, whose base is the number of crossings and whose height is the charge for each crossing. But the number of crossings has diminished, the benefit to bridge users has shrunk to the small triangular area at the top, and the total benefit has decreased by the area of the shaded triangle at the right. This triangle represents the net loss to society due to the faulty method of paying for the bridge. If, for example, the demand curve is a straight line, and if the owners set the toll so as to bring them a maximum return, the net loss of benefit is 25 per cent of the total.

These are the pertinent considerations if the bridge is already in existence, or its construction definitely decided. But if we examine the general question of the circumstances in which bridges ought to be built, a further inefficiency is disclosed in the scheme of paying for bridges out of tolls. For society, it is beneficial to build the bridge if the total area in the figure exceeds the interest, amortization, and maintenance costs. But if the bridge must be paid for by tolls, it will not be built unless it is expected that these costs will be exceeded by the rectangular area alone. This area cannot, for our example of a linear demand function, be greater than half the total. We may in this case say that the toll system has 75 per cent efficiency in use, but only 50 per cent efficiency in providing new bridges. In each case the efficiency will be further diminished by reason of the cost of collecting and accounting for the tolls.

The argument about bridges applies equally to railroads, except that in the latter case there is some slight additional cost resulting from an extra passenger or an extra shipment of freight. My weight is such that when I ride on the train, more coal has to be burned in the locomotive, and I wear down the station platform by walking across it. What is more serious, I may help to overcrowd the train, diminishing the comfort of other travelers and helping to create a situation in which additional trains should be run, but often are not. The trivial nature of the extra costs of marginal use of the railroads has from the

first been realized by the railroad managements themselves; indeed, it is implied in the amazingly complex rate structures they build up in the attempt to squeeze the last possible bit of revenue from freight and passenger traffic. If in a rational economic system the railroads were operated for the benefit of the people as a whole, it is plain that if people were to be induced by low rates to travel in one season rather than another, the season selected should be one in which travel would otherwise be light, leaving the cars nearly empty, and not a season in which they are normally overcrowded. Actually, our railroads run trains about the country in winter with few passengers, while crowding multitudes of travelers into their cars in summer. The rates are made high in winter, lower in summer, on the ground that the summer demand is more elastic than that of the winter travelers, who are usually on business rather than pleasure, and thus decide the question of a trip with less sensitiveness to the cost.

VI. COMPLEXITY OF ACTUAL RAILWAY RATES AND REMOTENESS FROM MARGINAL COST

The extreme and uneconomic complexity of railway freight and passenger rate structures is seldom realized by those not closely in touch with them. A few random examples will illustrate the remoteness of actual rates from what may be presumed to be marginal costs, which railway managements will find it profitable to cover even by the lowest rates. Prior to the last enforced reduction of American passenger rates the regular round-trip fare between New York City and Wilkesbarre, Pa. was $11.04. But at various times between 1932 and 1935 round-trip tickets good for limited periods were sold at $2.50, $6.00, $6.10, and $6.15. Between New York and Chicago the round-trip fare in the same period varied between $33 and $65 for identical accommodations. Between New York and Washington the ordinary round-trip fare was $18.00, but an "excursion rate" of $3.50 was applied spasmodically.

The lumber and logging activities of the country, which have been at a standstill for several years, are suffering from freight rates which in many important cases nearly equal, and even exceed, the mill price of the lumber. Thus from the large sawmills at and near Baker, Oregon, which produce lumber for the New York market, the freight amounts to $16.50 per thousand board feet. For No. 3 Common Ponderosa Pine, the grade shipped in largest quantities, the price of one-by-four inch boards ranged in the autumn of 1933 from $14.50 to $15.50 at the mill. Thus the New York wholesale buyer must pay more than double the mill price, solely on account of freight. The freight even to Chicago approximated the mill price. For No. 4 Common, also an important

grade, the price was $12.50 per thousand board feet at the mill, but the New York buyer had to pay $29.00. A few months earlier, the prices were about $8 per thousand board feet less than those just given, so that the railroads received far more than the mill operators. It is hard to escape the conclusion that these high freight rates interfered seriously with the sale of lumber.

One advantage of the system of charging only marginal cost would be a great simplification of the rate structure. This is a great desideratum. It must not be assumed too readily that every purchaser distributes his budget accurately to obtain the maximum of satisfactions, or the most efficient methods of production, when the determination of the optimum requires the study of an encyclopedic railroad tariff, together with complicated trial-and-error calculations. Neither, from the standpoint of a railroad, can it be assumed that the enormously complex rate differentials have been determined at all accurately for the purposes for which they were designed. These complicated rate structures further contravene the public interest in that they enhance artificially the advantages of large over small concerns. When immense calculations are required to determine the optimum combinations of transportation with other factors of production, the large concerns are in a distinctly better position to carry out the calculations and obtain the needed information.

VII. MARGINAL COST DEPENDS ON EXTENT OF UNUSED CAPACITY

In the determination of marginal cost there are, to be sure, certain complications. When a train is completely filled, and has all the cars it can haul, the marginal cost of carrying an extra passenger is the cost of running another train. On the other hand, in the more normal situation in which the equipment does not carry more than a small part of its capacity load, the marginal cost is virtually nothing. To avoid a sharp increase in rates at the time the train is filled, an averaging process is needed in the computation of rates, based on the probability of having to run an extra train. Further, in cases in which the available equipment is actually used to capacity, and it is not feasible or is of doubtful wisdom to increase the amount of equipment, something in the nature of a rental charge for the use of the facilities should, as indicated above, be levied to discriminate among different users in such a way that those willing to pay the most, and therefore in accordance with the usual assumptions deriving the most benefit, would be the ones obtaining the limited facilities which many desire. This rental charge for equipment, which for passenger travel would largely take the place of fares, should never be so high as to limit travel to fewer persons than can comfortably be accommodated, except for unpre-

dictable fluctuations. The proceeds from the charge could be added to the funds derived from income, inheritance, and land taxes, and used to pay a part of the overhead costs. But there should be no attempt to pay all the overhead from such rental charges alone.

Except in the most congested regions, there would, however, be no such charge for the use of track and stations until the volume of traffic comes to exceed enormously the current levels. An example is the great under-utilization of the expensive Pennsylvania Station, in New York City, whose capacity was demonstrated during the war by bringing into the city the trains of the Erie and the Baltimore and Ohio railroads. These trains are now required to stop on the New Jersey shore, constituting a wasteful nuisance which had existed before government operation, which was replaced by the more efficient procedure by the government, but which was resumed when the lines were handed back to their private owners.

VIII. THE ATTEMPT TO PAY FIXED COSTS FROM RATES AND PRICES CONTRIBUTES TO RIGIDITY AND SO TO INSTABILITY

One of the evil consequences of the attempt to pay overhead out of operating revenue is the instability which it contributes to the economic system as a whole. This is illustrated by the events leading to the depression. The immense and accelerating progress of science and technology led to the creation of new industries and the introduction of wonderfully efficient new methods. The savings from the new methods were so great that corporate profits and real incomes surged upwards. So large were the profits and so satisfactory the dividends that the operating officials of great industries did not feel under compulsion to push up the selling prices of their products to the levels corresponding to maximum monopoly profit. Because they kept their prices low, while paying relatively high wages, the physical volume of goods produced and transferred became enormous. The impulse to produce, with possibly some altruistic motives besides, tempered the desire for profits in many concerns. But under a profit system this could not last. As the prices of corporate shares rose, pressure developed to pay dividends equivalent to interest on the higher prices. This pressure would probably have led presently to gradual increases in the money prices of manufactured products, if the general level of prices had remained stationary. Such however was not the case. The general level of prices was declining.

And decline it must, according to the equation of exchange, when there was such a great new flood of goods to be sold. The vast increase in physical volume of goods, created by the new technology, called for a greater use of money, if the price level was to be maintained. This

need was met for a time by increases in bank loans and deposits, and in the velocity of circulation. But neither bank loans nor velocities could continue to increase as fast as goods, and prices had to fall. The fall was not uniform. Corporations under increasing pressure to cover their overhead and pay high dividends out of earnings were strongly averse to reducing the selling prices of their products, when these selling prices were already below the points which would yield maximum profit. For several years prior to the crash, the prices of manufactured products stuck fast, while the proportion of national expenditure paid for these products continued to increase. This left a shrinking volume of money payments to be made for the remaining commodities, and these, including particularly the agricultural, had to come down in price. If, as the general price level fell, railroad, utility, and manufacturing concerns had reduced their selling prices proportionately, the prosperity of the years 1922 to 1928 might have continued. But such reductions in selling prices were not possible when an increasing volume of overhead charges had to be paid out of earnings. The intensified efforts to do this resulted in a pushing up of "real" prices of manufactured products—that is, of the ratios of their prices to the general price level—and of "real" transportation rates. Indeed, with a rapidly falling general price level, railroad freight rates, measured in money, were actually increased in 1931. This increase of 15 per cent on a large range of commodities, like the subsequent increases in suburban commuters' passenger rates, was obtained on the ground that the railroads needed the money to cover their overhead costs, though their operating costs had declined. Of course the effect was to make the depression worse, by stopping traffic which would have flowed at the lower rates. On the theory that bond interest and other such items must be paid out of operating revenues, the railroads were "entitled" to the higher rates, for their business had fallen off. But economic equilibrium calls for a rising rather than a declining supply curve; if demand falls off, the offer price must be reduced in order to have the offered services taken. This antithesis of rising railway rates, when general prices and the ability to pay are falling, well illustrates the disequilibrating consequences of the idea that overhead costs must be paid from operating revenues. There now seems to be a possibility of a repetition of the disastrous 15 per cent freight-rate increase in a time of decline.[10]

This explanation of the contrast of the prosperity of 1928 with the cessation of production in the following years rests upon the contrast of the system of prices which results from the whole-hearted devotion

[10] Since this was written the Interstate Commerce Commission has allowed a part of this proposed increase and postponed consideration of a request for a passenger fare rise.

of different concerns to their own respective profits, with the system of prices best for the economic organism as a whole. Under free competition, with no overhead, these two systems of prices tend to become identical. Where there are overhead costs, competition of the ideally free type is not permanently possible. Monopoly prices develop; and a system of monopoly prices is not a system which can serve human needs with maximum advantage.

IX. CRITERION AS TO WHAT INVESTMENTS ARE SOCIALLY WORTH WHILE

When a decision whether or not to construct a railway is left to the profit motive of private investors, the criterion used is that the total revenue $\sum p_i q_i$, being the sum of the products of the rates for the various services by the quantities sold, shall exceed the sum of operating costs and carrying charges on the cost of the enterprise. If no one thinks that there will be a positive excess of revenue, the construction will not be undertaken. We have seen in Section V that this rule is, from the standpoint of the general welfare, excessively conservative. What, then, should society adopt to replace it?

A less conservative criterion than that of a sufficient revenue for total costs is that *if some distribution of the burden is possible such that everyone concerned is better off than without the new investment, then there is a prima facie case for making the investment.* This leaves aside the question whether such a distribution is *practicable.* It may often be good social policy to undertake new enterprises even though some persons are put in a worse position than before, provided that the benefits to others are sufficiently great and widespread. It is on this ground that new inventions are permitted to crowd out less efficient industries. To hold otherwise would be to take the side of the hand weavers who tried to wreck the power looms that threatened their employment. But the rule must not be applied too harshly. Where losses involve serious hardship to individuals, there must be compensation, or at least relief to cover subsistence. Where there are many improvements, the law of averages may be trusted to equalize the benefits to some extent, but never completely. It will always be necessary to provide for those individuals upon whom progress inflicts special hardship; if it were not possible to do this, we should have to reconcile ourselves to greater delays in the progress of industrial efficiency.

Subject to this qualification of avoiding excessive hardship to individuals, we may adopt the criterion stated. In applying it there will be the problem of selecting a limited number of proposed investments, corresponding to the available capital, from among a larger number of possibilities. The optimum solution corresponds to application of our

criterion to discriminate between each pair of combinations. The total amount of calculation and exercise of judgment required will not, however, be so great as might be suggested by the number of pairs of combinations, which is immense. Numerous means are available to shorten this labor. One of these is by the application of the line integral (3), namely

$$w = \int \sum h_i dq_i,$$

which provides a measure of value corresponding to the sum of consumers' and producers' surpluses. The part of w constituting the generalized consumers' surplus is (2); the validity of this line integral as a measure of an individual's increment of satisfaction corresponding to sufficiently small changes in the q's may be seen merely by replacing p_i in (13) by f_i, and noticing that for small changes the marginal utility of money λ changes little, so that f_i is very nearly proportional to the derivative of the utility function Φ. Hence the increment in Φ is proportional to the sum of the integrals of the f's, apart from terms of higher order; and the factor of proportionality λ is such as to measure this increment in money so as to be comparable to an increase in income. Similar considerations apply to the part of w corresponding to producers' surplus.

Defenders of the current theory that the overhead costs of an industry must be met out of the sale of its products or services hold that this is necessary in order to find out whether the creation of the industry was a wise social policy. Nothing could be more absurd. Whether it was wise for the government to subsidize and its backers to construct the Union Pacific Railroad after the Civil War is an interesting historical question which would make a good subject for a dissertation, but it would be better, if necessary, to leave it unsolved than to ruin the country the Union Pacific was designed to serve by charging enormous freight rates and claiming that their sum constitutes a measure of the value to the country of the investment. Such an experimental solution of a historical question is too costly. In addition, it is as likely as not to give the wrong answer. The sum of the freight and passenger rates received, minus operating costs, is not the line integral $w=\int\sum h_i dq_i$, which with some accuracy measures the value to society of the investment, but is more closely related to the misleading measure of value $\sum p_i q_i$. In other words, the revenue resembles the area of the rectangle in Figure 2, while the possible benefit corresponds to the much larger triangular area. The revenue is the thing that appeals to an investor bent on his own profit, but as a cri-

terion of whether construction ought in the public interest to be undertaken, it is biased in the direction of being too conservative.

Regardless of their own history, the fact is that we now have the railroads, and in the main are likely to have them with us for a considerable time in the future. It will be better to operate the railroads for the benefit of living human beings, while letting dead men and dead investments rest quietly in their graves, and to establish a system of rates and services calculated to assure the most efficient operation. When the question arises of building new railroads, or new major industries of any kind, or of scrapping the old, we shall face, not a historical, but a mathematical and economic problem. The question then will be whether the aggregate of the generalized surpluses of the form (3) is likely to be great enough to cover the anticipated cost of the new investment. This will call for a study of demand and cost functions by economists, statisticians, and engineers, and perhaps for a certain amount of large-scale experimentation for the sake of gaining information about these functions. The amount of such experiment and research which could easily be paid for out of the savings resulting from operation of industry in the public interest is very large indeed. Perhaps this is the way in which we shall ultimately get the materials for a scientific economics.

Columbia University
New York, N. Y.

MANAGING EDITOR'S NOTE: Professor Frisch has written a brief criticism of Professor Hotelling's argument, but because of limitations of space it has had to be held over for publication in a later issue.—D. H. L.

DIFFERENTIAL EFFECT IN THE BUTTER MARKET

By Irma Hilfer

I. INTRODUCTION

In a recent study of the Danish butter market,[1] the regression equations of the wholesale prices of Danish and New Zealand butter, on the sales of each of three types, for the period from February, 1930 through October, 1936, had the form:[2]

$$y_1 = -1.892x_1 - .786x_2 - .612x_3 + 1{,}268.60$$
$$y_2 = -1.549x_1 - .988x_2 - .516x_3 + 1{,}165.61$$

The prices of Danish and New Zealand butter, with several adjustments made, are represented by y_1 and y_2 respectively; and the adjusted sales of Danish, New Zealand and Australian, and all other types of butter, by x_1, x_2, and x_3.

Since all the sales are measured in the same units, the equations seem to show that the sale of Danish butter has a greater effect on the prices of both Danish and New Zealand butter than have the sales of any other type of butter.

Whether or not this result has arisen from sampling errors can be determined only by some statistical test of the variation among the regression coefficients.

II. A TEST OF DIFFERENTIAL EFFECT

Determining whether the variation among a set of p regression coefficients is greater than that caused by chance presents an interesting problem. The two methods which first suggest themselves are testing the differences between the largest coefficient and each of the others by means of Student's "t" test, and testing the ratio of the sum of the squares of the regression coefficients to some theoretical value for it, by means of the analysis of variance. The first method has a definite bias, since it selects the largest regression coefficient; the second is invalid. The observations whose variance is being tested are assumed independent; while the covariance of regression coefficients b_i and b_j is c_{ij}, the cofactor of the term in the ith row and the jth column of $|a_{ij}|$, the determinant of sums of squares and sums of products of the independent variables, divided by $|a_{ij}|$ itself, multiplied by σ^2, the true variance of the residuals of the y's. Therefore, unless every non-principal minor of $|a_{ij}|$ is zero, the second method is invalid here.

In more rigorous form, the problem is to determine, when given a

[1] J. Pedersen, *En Analyse af det Engelske Smörmarked, i Perioden 1923–1936*, Aarhus Universitets Okonomiske Institut, 1937.

[2] *Ibid.*, pp. 39 and 41. (Slight changes were made in the notation.)

regression equation of y on $x_1, \cdots x_p$, whether there is significant evidence that the hypothesis that the distribution of y is of the form:

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2\sigma^2)[y-\beta(x_1+\cdots+x_p)]^2} dy$$

is false. The number of observations is N; and all the variates are assumed to be deviations from their respective means, without any loss of generality. Let the independent variates, $x_1, \cdots, x_p$, be transformed so that:

$$x_1' = k(x_1 + \cdots + x_p),$$
$$x_i' = \sum m_{ij}x_j, \qquad (i = 2, \cdots, p),$$

and

$$Sx_i'x_j' = \delta_{ij}.$$

The symbol δ_{ij} is the Kronecker delta, when $i=j$, $\delta_{ij}=1$; $i\neq j$, $\delta_{ij}=0$. Throughout the paper, $\sum$ denotes summation from 1 to p, and S summation from 1 to N.

The regression equation is transformed to

$$y = b_1'x_1' + \cdots + b_p'x_p';$$

and the determinant $|a_{ij}|$ to the identical determinant

$$\begin{vmatrix} 1 & 0 \cdots & & 0 \\ 0 & 1 \cdots & & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 \cdots & 0 & 1 \end{vmatrix}$$

in which $c_{ij}=\delta_{ij}$.

If the hypothesis is true,

$$Eb_1' = \beta,$$
$$Eb_i' = 0, \qquad (i = 2, \cdots, p),$$
$$Eb_i'b_j' - (Eb_i')(Eb_j') = \delta_{ij}\sigma^2.$$

The new regression coefficients are, then, normally and independently distributed. The sum of the squares of the last $p-1$ of them, $\sum b_i'^2 - b_1'^2$, is an adequate test of deviation from hypothesis, and the coefficient of differential effect should be proportional to it. Since $Eb_i'^2=\sigma^2$, $(i=2, \cdots, p)$, $[\sum b_i'^2 - b_1'^2]/(p-1)$ is an estimate of the variance of the residuals. Then

$$D = \frac{N-p-1}{p-1} \frac{\sum b_i'^2 - b_1'^2}{S(y - b_1x_1 - \cdots - b_px_p)^2},$$

the ratio of this estimate of the variance to the usual sample estimate,

has the distribution of Snedecor's F, the ratio of two variances, with $p-1$ and $N-p-1$ degrees of freedom.

For practical reasons, it is necessary to find D as a function of the original variates. To find $\sum b_i'^2$, we note from least-squares theory, that

$$b_i' = \sum_k c_{ik}' l_k',$$

where

$$l_k' = Syx_k'.$$

Since

$$c_{ik}' = \delta_{ik}, \text{ here,}$$

$$b_i' = l_i',$$

and

$$\sum b_i'^2 = \sum b_i' l_i'.$$

The multiple-correlation coefficient, R, which Dr. Hotelling has shown[3] is invariant under internal linear transformations of the x's, is given by:

$$R^2 = \sum b_i' l_i' / Sy^2.$$

Then

$$\sum b_i'^2 = (Sy^2)R^2,$$

$$b_1' = Syx_1' = kSy(x_1 + \cdots + x_p),$$

where $k^2 = 1/S(x_1 + \cdots + x_p)^2$, because of the conditions of the transformation. Then

$$b_1'^2 = Sy^2 r^2_{y\Sigma x},$$

and

$$\sum b_i'^2 - b_1'^2 = Sy^2(R^2 - r^2_{y\Sigma x}).$$

From another formula for the multiple-correlation coefficient, we find,

$$S(y - b_1x_1 - \cdots - b_px_p)^2 = Sy^2(1 - R^2).$$

Substituting,

$$D = \frac{(N - p - 1)}{(p - 1)} \frac{(R^2 - r^2_{y\Sigma x})}{(1 - R^2)} \cdot$$

The formula is valid when the variates are measured from some origin other than the mean, and the regression equation contains a constant term; since a simple transformation of the variates will reduce the equation to the type studied, without affecting the correlations.

The above form for D shows its logical basis, since, when the hypothesis is true, there is no more correlation of y with the independent

[3] H. Hotelling, "Relations between Two Sets of Variates," *Biometrika*, Vol. 28, Dec., 1936, pp. 321–377.

variates taken separately than of y with an equally weighted linear function of them. For computational purposes, however, we can substitute for the correlation of y with the total of the independent variates, since

$$r_{y\Sigma x} = \frac{\sum Sy x_i}{\sum_j \sum_i S x_i x_j} .$$

The formula for D could have been derived by the use of the analysis of variance. If the sum of squares is apportioned to the residuals, the first transformed variate, x_1', and the last $p-1$ transformed variates, the first portion is $Sy^2\,(1-R^2)$, the second, $Sy^2 r^2{}_{\Sigma yx}$, and the last, $Sy^2(R^2-r^2{}_{y\Sigma x})$, with $N-p-1$, 1, and $p-1$ degrees of freedom respectively. The division of the sum of squares attributed to the independent variates is valid because of the orthogonality of the transformation. A comparison of the estimates of variance obtained from the sum of squares attributed to the $p-1$ transformed x's and from the residuals gives the statistic D.

III. APPLICATION OF THE TEST

Before using Pedersen's equations to determine whether the sale of any type of butter has a significantly greater effect on the price of either Danish or New Zealand butter, it seemed wise to examine the method used in obtaining his results. The report states that in the equations:

$$y_1 = -\,1.892x_1 - .786x_2 - .612x_3 + 1{,}268.60$$

$$y_2 = -\,1.549x_1 - .988x_2 - .516x_3 + 1{,}165.61,$$

the significance of the variates is as follows:

y_1 is the wholesale price of Danish butter in England, in shillings per hundredweight, divided by an index of wages and salaries in England, and multiplied by ten thousand;

y_2 is the wholesale price of New Zealand butter, similarly deflated, and expressed in the same units;

x_1 is England's imports of Danish butter, expressed in thousands of hundredweights;

x_2 is England's imports of New Zealand and Australian butter, plus the amount of butter taken out of cold storage, expressed in the same units;

x_3 is England's imports of butter from all other countries, expressed in the same units.

The period considered was from February, 1930 through October, 1936, with monthly data. The adjustment for the change in the amount of butter stored was made on the New Zealand and Australian butter only since it was found that most of the stored butter was of this type.

TABLE 1

ENGLAND'S IMPORTS OF BUTTER, BY EXPORTING COUNTRY, 1930–1936*
Unit: one thousand hundredweight.

	Denmark	Soviet Union	Holland	Irish Free State	Australia	New Zealand	Other Countries
1930							
March	181	—	5	6	120	141	114
April	175	4	10	18	69	215	93
May	187	18	14	48	100	133	109
June	260	16	14	97	68	76	152
July	213	30	8	99	21	90	80
August	199	9	9	82	23	110	112
September	197	45	6	75	28	71	75
October	212	20	7	58	54	74	90
November	163	13	4	24	69	133	55
December	204	5	5	5	166	181	95
1931							
January	158	—	4	3	191	231	88
February	154	—	3	2	123	230	85
March	181	—	5	2	123	163	89
April	214	2	8	6	153	182	106
May	216	31	12	35	148	170	111
June	252	37	16	67	148	114	123
July	244	76	11	84	79	156	158
August	229	74	10	68	65	115	97
September	219	51	10	46	68	98	105
October	225	63	7	44	114	88	100
November	190	58	4	21	115	131	78
December	185	12	4	4	231	256	92
1932							
January	174	3	2	2	183	256	91
February	179	—	2	2	233	207	105
March	146	—	2	1	177	259	71
April	183	13	3	2	112	186	100
May	256	13	3	30	119	174	100
June	286	29	3	65	122	141	137
July	268	75	8	66	89	149	149
August	260	52	8	37	114	102	120
September	220	72	4	57	190	155	103
October	214	59	4	39	108	61	76
November	202	6	5	13	169	226	54
December	197	—	2	1	313	214	57
1933							
January	173	5	2	1	241	253	70
February	172	3	4	1	184	191	72
March	201	17	6	14	233	277	81
April	225	33	8	11	188	220	80
May	239	29	34	46	174	173	72
June	266	15	65	82	80	257	88

* Source: Pedersen, *op. cit.*, pp. 95–97.

TABLE I (*Continued*)

	Denmark	Soviet Union	Holland	Irish Free State	Australia	New Zealand	Other Countries
1933							
July	244	93	14	94	46	127	114
August	215	152	3	83	44	226	96
September	224	76	2	45	31	158	107
October	194	75	2	1	69	152	107
November	178	30	3	1	131	229	69
December	189	34	3	1	273	249	68
1934							
January	205	15	4	1	272	322	81
February	177	12	3	1	236	276	85
March	212	14	11	4	250	362	87
April	209	41	36	19	180	186	64
May	264	28	62	58	200	287	94
June	287	71	64	95	156	170	163
July	246	71	45	95	153	196	159
August	213	130	30	73	87	137	136
September	192	4	28	57	53	122	107
October	181	113	13	53	83	209	77
November	143	19	8	12	189	167	54
December	157	—	7	2	245	243	53
1935							
January	152	—	6	2	301	267	50
February	147	—	9	4	156	290	66
March	173	—	26	1	362	263	69
April	187	36	40	10	317	236	66
May	204	71	66	68	236	217	97
June	235	65	66	99	107	161	118
July	233	84	56	121	93	167	164
August	203	106	65	84	57	195	149
September	165	77	39	48	36	179	126
October	157	27	40	37	86	187	181
November	163	30	18	13	146	154	72
December	166	11	32	3	216	320	63
1936							
January	157	—	27	2	250	271	77
February	167	—	18	2	182	258	95
March	152	—	42	2	206	255	92
April	176	—	66	11	237	280	91
May	202	29	92	54	166	250	120
June	198	34	91	89	75	204	186
July	206	83	83	89	107	175	190
August	188	28	88	92	50	148	164
September	190	95	74	63	85	216	148
October	196	79	63	50	96	173	120
November	159	44	54	28	124	206	98

Since the cold-storage figures were given fortnightly, it was necessary to interpolate linearly for an estimate of the amount in storage at the end of each month, and then to make the necessary adjustments. The author felt that secular trend was not a relevant factor here, and did not consider it in his calculations. The price, import, and cold-storage figures were obtained from the newspaper, *Landbrugraadets Meddelser;* and the index of wages and salaries was interpolated from one given by Colin Clark in *National Income and Outlay*. The series are shown in Tables 1, 2, 3, and 4.

The assumptions stated could be justified on the basis of the author's knowledge of the subject, and may be accepted; but it was found that, in addition, all the variates were smoothed by a three months' moving average before any computations were made. Smoothing a variate x in this manner is equivalent to transforming all the observations, so that the ith pseudo-observation x_i', is equal to $(x_{i-1}+x_i+x_{i+1})/3$. Then the first serial correlation, or the correlation between consecutive pseudo-observations, is approximately

$$(2Ex_i^2 + 4Ex_ix_{i+1} + 2Ex_ix_{i+2} + Ex_ix_{i+3})/9Ex_i^2.$$

TABLE 2

TONS OF BUTTER IN COLD STORAGE*

1930		1931		1932		1933	
Date	Amount	Date	Amount	Date	Amount	Date	Amount
1/11	8,256	1/10	5,972	1/ 9	9,246	1/ 7	14,973
1/25	11,783	1/24	7,653	1/23	10,242	1/21	16,210
2/ 8	12,956	2/ 7	7,125	2/ 6	9,463	2/ 4	19,356
2/22	11,727	2/21	7,636	2/20	7,909	2/18	20,823
3/ 8	15,273	3/ 7	6,774	3/ 5	9,096	3/ 4	20,680
3/22	15,800	3/21	7,384	3/19	9,700	3/18	20,589
4/ 5	16,812	4/ 4	6,381	4/ 2	11,369	4/ 1	22,673
4/19	18,479	4/18	7,579	4/16	12,300	4/15	23,932
5/ 3	21,655	5/ 2	9,222	4/30	13,438	4/29	24,950
5/17	22,989	5/16	12,324	5/14	11,696	5/13	25,087
5/31	26,178	5/30	14,179	5/28	13,154	5/27	29,306
6/14	28,723	6/13	15,763	6/11	16,400	6/10	32,072
6/28	30,333	6/27	18,194	6/25	18,649	6/24	33,180
7/12	30,185	7/11	20,261	7/ 9	20,917	7/ 8	32,826
7/26	26,987	7/25	23,445	7/23	22,256	7/22	33,226
8/ 9	25,514	8/ 8	26,754	8/ 6	22,098	8/ 5	32,522
8/23	25,460	8/22	24,931	8/20	21,842	8/19	31,455
9/ 6	24,581	9/ 5	23,381	9/ 3	19,691	9/ 2	27,949
9/20	22,545	9/19	23,191	9/17	18,182	9/16	24,473
10/ 4	21,046	10/ 3	19,995	10/ 1	18,465	9/30	22,346
10/18	18,882	10/17	18,421	10/15	19,109	10/14	20,215
11/ 1	15,587	10/31	15,035	10/29	15,603	10/28	16,007
11/15	11,134	11/14	14,693	11/12	12,008	11/11	14,354
11/29	8,135	11/28	12,228	11/26	9,954	11/25	11,777
12/13	7,011	12/12	9,891	12/10	12,132	12/ 9	11,175
12/27	6,404	12/26	11,019	12/24	14,899	12/23	10,324

* Source: Pedersen, *op. cit.*, p. 100.

TABLE 2 (*Continued*)

1934		1935		1936	
Date	Amount	Date	Amount	Date	Amount
1/ 6	13,001	1/ 5	12,901	1/ 4	8,909
1/20	15,088	1/19	11,946	1/18	9,522
2/ 3	16,040	2/ 2	9,844	2/ 1	11,808
2/17	19,058	2/16	11,845	2/15	11,316
3/ 3	20,211	3/ 2	10,697	2/29	10,933
3/17	22,242	3/16	11,472	3/14	10,083
3/31	25,638	3/30	12,032	3/28	9,760
4/14	24,487	4/13	12,594	4/11	10,199
4/28	27,881	4/27	14,134	4/25	10,147
5/12	28,969	5/11	20,984	5/ 9	13,587
5/26	29,818	5/25	23,377	5/23	16,310
6/ 9	37,511	6/ 8	25,226	6/ 6	18,141
6/23	41,612	6/22	26,619	6/20	21,439
7/ 7	44,877	7/ 6	31,330	7/ 4	22,492
7/21	47,554	7/20	31,888	7/18	24,985
8/ 4	47,884	8/ 3	30,134	8/ 1	28,000
8/18	47,175	8/17	29,953	8/15	24,613
9/ 1	45,406	8/31	27,049	8/29	25,292
9/15	42,299	9/14	22,122	9/12	26,308
9/29	35,461	9/28	21,114	9/26	25,691
10/13	30,948	10/12	17,625	10/10	25,066
10/27	27,091	10/26	15,476	10/24	23,954
11/10	22,292	11/ 9	13,615	11/ 7	21,588
11/24	18,846	11/23	11,258	11/21	21,108
12/ 8	16,888	12/ 7	9,989	12/ 5	19,887
12/22	15,559	12/21	8,578	12/29	18,145

TABLE 3

INDEX OF ENGLISH WAGES AND SALARIES*

	1930	*1931*	*1932*	*1933*	*1934*	*1935*	*1936*
January	2,400	2,278	2,226	2,231	2,323	2,413	2,509
February	2,386	2,272	2,224	2,237	2,331	2,421	2,518
March	2,373	2,267	2,221	2,244	2,339	2,429	2,528
April	2,360	2,262	2,218	2,252	2,347	2,437	2,538
May	2,347	2,258	2,216	2,258	2,353	2,444	2,548
June	2,336	2,254	2,214	2,266	2,360	2,453	2,558
July	2,325	2,249	2,212	2,275	2,368	2,460	2,566
August	2,316	2,245	2,213	2,283	2,376	2,469	2,576
September	2,307	2,241	2,215	2,292	2,383	2,477	2,585
October	2,298	2,238	2,217	2,300	2,390	2,485	2,595
November	2,290	2,234	2,220	2,307	2,398	2,493	2,605
December	2,284	2,230	2,225	2,315	2,404	2,501	2,615

* Source: Pedersen, *op. cit.*, p. 102.

Therefore, unless certain narrow and infrequently met conditions on the serial correlations of the original observations are true, the new observations are not independent; and least-squares equations on them do not have the accepted meaning. The variations from the trend

TABLE 4

WHOLESALE PRICES OF DANISH AND NEW ZEALAND BUTTER IN ENGLAND, 1930–1936*

Unit: Shillings per hundredweight.

	1930		1931		1932		1933	
	Danish	New Zealand	Danish	New Zealand	Danish	New Zealand	Danish	New Zealand
January	173	158	136	117	125	100	111	83
February	176	152	149	122	143	107	107	79
March	161	140	137	119	127	110	98	75
April	140	129	124	111	117	107	94	69
May	134	131	120	111	104	97	94	79
June	140	135	119	112	102	98	91	81
July	151	138	119	114	112	105	97	81
August	149	136	125	115	113	110	104	90
September	151	130	131	115	124	115	116	102
October	151	120	134	121	119	113	113	101
November	142	112	132	114	124	96	119	87
December	139	113	137	104	122	85	110	71

	1934		1935		1936	
	Danish	New Zealand	Danish	New Zealand	Danish	New Zealand
January	90	67	118	82	120	95
February	92	70	117	89	129	94
March	94	72	106	76	123	85
April	83	71	98	78	106	88
May	86	77	93	80	104	95
June	86	78	101	88	116	108
July	89	75	103	91	125	115
August	107	82	111	95	128	120
September	107	78	127	112	125	108
October	112	71	132	124	123	100
November	121	75	125	108	123	109
December	124	73	128	91	115	101

* Source: Pedersen, *op. cit.*, p. 99.

found by the moving average contain a great deal of relevant information which is lost by smoothing.

It was decided to adjust the series as Pedersen did, except for the smoothing, and to recompute the equations, in order to test for differential effect. The variates used are shown in Table 5. Since the computed values of the prices obtained in the published study seemed highly correlated with the prices two months later, it was decided to obtain the regression equations with quantity lagging two months be-

TABLE 5

UNSMOOTHED VARIATES USED IN COMPUTATION

y_1 = 10,000 × the price of Danish Butter per cwt. divided by the Index of Wages and Salaries.
y_2 = 10,000 × the price of New Zealand Butter per cwt. divided by the Index of Wages and Salaries.
x_1 = Danish Butter imported in thousands of hundredweights.
x_2 = New Zealand and Australian Butter imported, plus the amount removed from cold storage during the month, in thousands of hundredweights.
x_3 = Butter imported from Denmark, Soviet, Holland, Irish Free State, and all other countries during the months, in thousands of hundredweights.

Date	y_1 (1)	y_2 (2)	x_1 (3)	x_2 (4)	x_3 (5)
1930					
January	720.8	658.3	—	—	—
February	737.6	637.0	166	211	128
March	678.5	590.0	181	197	125
April	593.2	546.6	175	194	125
May	570.9	558.2	187	129	189
June	599.3	577.9	260	61	279
July	649.5	593.5	213	188	217
August	643.4	587.2	199	163	212
September	654.5	563.5	197	169	201
October	657.1	522.2	212	241	175
November	620.1	489.1	163	357	96
December	608.6	494.7	204	383	110
1931					
January	597.0	513.6	158	400	95
February	655.8	537.0	154	349	90
March	604.3	524.9	181	297	96
April	548.2	490.7	214	289	122
May	531.4	491.6	216	212	189
June	528.0	496.6	252	175	243
July	529.1	506.9	244	110	329
August	556.8	512.2	229	199	249
September	584.6	513.2	219	231	212
October	598.7	540.7	225	315	214
November	590.9	510.3	190	309	161
December	614.3	466.4	185	517	112
1932					
January	561.5	449.2	174	451	98
February	643.0	481.1	179	463	109
March	571.8	495.3	146	387	74
April	527.5	482.4	183	252	118
May	469.3	437.7	256	285	146
June	460.7	442.6	286	151	234
July	506.3	474.7	268	184	298
August	510.6	497.1	260	256	217
September	559.8	519.2	220	279	236
October	536.8	509.7	214	236	178
November	558.6	432.4	202	485	78
December	548.3	382.0	197	517	60

TABLE 5 (*Continued*)

Date	y_1	y_2	x_1	x_2	x_3
	(1)	(2)	(3)	(4)	(5)
1933					
January	497.5	372.0	173	424	78
February	478.3	353.2	172	328	80
March	436.7	334.2	201	474	118
April	417.4	306.4	225	359	132
May	416.3	349.9	239	276	180
June	401.6	357.5	266	247	250
July	426.4	356.0	244	178	315
August	455.5	394.2	215	356	334
September	506.1	445.0	224	311	230
October	491.3	439.1	194	355	185
November	515.8	377.1	178	442	103
December	475.2	306.7	189	516	106
1934					
January	387.4	288.4	205	514	101
February	394.7	300.3	177	430	101
March	401.9	307.8	212	498	116
April	353.6	302.5	209	318	160
May	365.5	327.2	264	396	242
June	364.4	330.5	287	113	393
July	375.8	316.7	246	257	370
August	450.3	345.1	213	269	369
September	449.0	327.3	192	382	196
October	468.6	297.1	181	480	256
November	504.6	312.8	143	513	93
December	515.8	303.7	157	569	62
1935					
January	489.0	339.8	152	642	58
February	483.3	367.6	147	432	79
March	436.4	312.9	173	601	96
April	402.1	320.1	187	482	152
May	380.5	327.3	204	282	302
June	411.7	358.7	235	165	348
July	418.7	369.9	233	236	425
August	449.6	384.8	203	321	404
September	512.7	452.2	165	344	290
October	531.2	499.0	157	389	285
November	501.4	433.2	163	384	133
December	511.8	363.9	166	572	109
1936					
January	478.3	378.6	157	465	106
February	512.3	373.3	167	454	115
March	486.6	336.2	152	483	136
April	417.7	346.7	176	486	168
May	408.2	372.8	202	296	295
June	453.5	422.2	198	182	400
July	487.1	448.2	206	170	445
August	496.9	465.8	188	245	352
September	483.6	417.8	190	299	380
October	474.0	385.4	196	324	312
November	472.2	418.4	159	379	224

hind price, as well. In order to keep within the period for which the figures were complete, the quantities were taken from March, 1930 through November, 1936, in both sets of equations, and the prices for corresponding periods. The equations with no lag are not strictly comparable with those published, but would change very slightly if the period studied were February, 1930 through October, 1936.

The sample variances and covariances found from the data are shown in the table:—

	y_1 (No lag)	y_2 (No lag)	y_1 (2 months' lag)	y_2 (2 months' lag)	x_1	x_2	x_3
y_1 (no lag)	6,437.15						
y_2 (no lag)		7,481.34					
y_1 (2 months' lag)			7,614.74				
y_2 (2 months' lag)				8,673.65			
x_1	− 756.26	+ 162.36	− 801.13	− 132.80	1,173.17		
x_2	−1,757.41	−6,288.26	−1,097.57	−3,954.75	−2,786.87	16,683.98	
x_3	−2,642.02	+ 160.20	−4,847.64	−2,429.25	+1,849.44	−8,273.04	10,509.39

The use of the inverse matrix of the variances and covariances of the x's in solving the normal equations was very economical here. The elements of the matrix are:

$$\begin{array}{lll} +\ .001484670 & & \\ +\ .000194280 & +\ .000123737 & \\ -\ .000108349 & +\ .000063217 & +\ .000163983. \end{array}$$

The regression equations, where no lag was used, are:

$$y_1 = -\ 1.1780x_1 - .5314x_2 - .4624x_3 + 1009.19,$$

$$y_2 = -\ \ .9980x_1 - .7364x_2 - .3888x_3 + \ \ 945.30;$$

and, where a two months' lag was used:

$$y_1 = -\ .8774x_1 - .5979x_2 - .7775x_3 + 1039.21,$$

$$y_2 = -\ .7023x_1 - .6687x_2 - .6340x_3 + \ \ 917.44.$$

The statistics of differential effect, and the correlations used in computing them, are:

	y_1 (no lag)	y_2 (no lag)	y_1 (2 months' lag)	y_2 (2 months' lag)
R^2	.4873	.5890	.6735	.4932
$r^2_{y\Sigma x}$	.4152	.4783	.6010	.4921
$D = 38.5 \dfrac{R^2 - r^2_{y\Sigma x}}{1 - R^2}$	4.243	10.37	8.551	.0770
$\frac{1}{2}$ Log $D = z$	.8531	1.806	1.267	−2.6
Significant or not significant	Significant	Significant	Significant	Not significant

While the butter prices are highly correlated with the quantities sold, Pedersen's incorrect multiple correlations, which were greater than .8, are reduced. The increase in the correlation of the price of Danish butter with the sales effected by the two months' lag and the decrease in the New Zealand butter correlation are interesting. The true lag may be somewhat less than two months, with one set of correlations decreasing from the maximum more sharply than the other; or the more frequent cold storage of New Zealand and Australian butter may influence the market. Since the one per cent point for z, is .8025, marked differential effect of the sales of the different types of butter on the price of Danish butter has been found, whether or not the equation is formed with quantity lagging behind price, and on the price of New Zealand butter, only when no lag is used.

IV. GENERALIZATION TO A SET OF DEPENDENT VARIATES

It seems natural to consider the case of testing, when given a set of m simultaneous equations of the type

$$y_j = \sum_i b_{ji} x_i, \qquad (j = 1, 2, \cdots, m), \ (i = 1, 2, \cdots, p),$$

whether the x's differ significantly in their effect on the set of y's; and to generalize the normal frequency distribution which states the hypothesis in Section II to the multivariate normal distribution of the residuals:

$$y_j - \beta_j(x_1 + \cdots + x_p).$$

If the x's are transformed as before, the appropriate test of differential effect seems to be proportional to the mth order determinant,

$$\left| \left(\sum b_{ji}' b_{ki}' - b_{j1}' b_{k1}' \right) \right|,$$

if the number of independent variates is greater than the number of dependent variates; but recent work on multivariate analysis shows that this function cannot be used in an adequate statistic.

The determinant above can be obtained by multiplying the matrix

$$\left\| \begin{matrix} b_{12}' & \cdots & b_{1p}' \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ b_{m2}' & \cdots & b_{mp}' \end{matrix} \right\|$$

by itself, row by row; and, as Dr. Hotelling has shown[4] is proportional

[4] *Loc. cit.*

to the vector correlation coefficient, q^2. In terms of $r_1, \cdots, r_{p-1}$, the canonical correlations, or the correlations between corresponding linear functions of the variates of each set, selected to have maximum correlation, $q^2 = r_1^2 r_2^2 \cdots r_{p-1}^2$.

If the hypothesis is satisfied, the matrix of correlations of the my's with the $p-1$ transformed x's is the zero matrix; and, therefore,[5] all the canonical correlations will vanish. Since the mth order determinant above will vanish when any one of the canonical correlations vanishes, it cannot be used in a test of the hypothesis, unless $m=1$.

A satisfactory test of differential effect for two sets of variates could be made by using z, the generalization of $1-R,^2$ for the my's and the $p-1$ transformed x's;[6] but no distribution of this statistic for the case where one set of variates is normally distributed and the other is fixed has been found yet.

V. SUMMARY

The conclusions found from this study, are:

1. While the prices of different types of butter are very closely related to the sale of different types of butter, the high degree of relationship derived in Jörgen Pedersen's study is erroneous, because of an error in the method used to find the regression equations.

2. After testing by a method developed here, it has been found that the sales of three different types of butter differ significantly in their effect on the prices.

3. The sales of the three types of butter are more closely related to the wholesale price of Danish butter two months previously than to the current wholesale price; but the opposite is true for the price of New Zealand butter.

While a generalization of D to a set of dependent variates is invalid, another type of generalization can be made. If it is desired to test whether the true value of a regression equation,

$$y = b_1 x_1 + \cdots + b_p x_p,$$

differs from some theoretical equation,

$$y = \beta_1 x_1 + \cdots + \beta_p x_p,$$

one can test the statistic

[5] *Ibid.*

[6] S. S. Wilks, "Certain Generalizations in the Analysis of Variance," *Biometrika*, Vol. 24, 1932, pp. 471–494.

$$D' = \frac{N - p - 1}{p - 1} \frac{R^2 - r^2_{y\Sigma\beta x}}{1 - R^2}$$

in the same manner as was used in testing D.

Columbia University
New York, N. Y.

REFERENCES

FISHER, R. A. *Statistical Methods for Research Workers*, London, 1936.

HOTELLING, H. "Relations between Two Sets of Variates," *Biometrika*, Vol. 28, December, 1936, pp. 321–377.

PEDERSEN, J. *En Analyse af det Engelske Smormarked, i Perioden 1923–1936*, Aarhus Universitets Okonomiske Institut, Aarhus, 1937.

WILKS, S. S. "Certain Generalizations in the Analysis of Variance," *Biometrika*, Vol. 24, 1932, pp. 471–494.

THE DEFINITION OF "EQUAL WELL-BEING" IN FRISCH'S DOUBLE EXPENDITURE METHOD

BY HORST MENDERSHAUSEN

IN THE PUBLICATIONS concerning his double-expenditure method[1] Professor Frisch gave a criterion of equal well-being, which gives a sufficiently close approximation under certain assumptions. The following is a suggestion of a simple interpretation of this criterion.

Given are two social groups (e.g., groups of families) A and B with the same want constitution and living in structurally equal markets (i.e., the same kinds of goods are available). The groups find themselves in two different price situations, viz., A in 0 and B in 1. They compose their quantity budgets[2] in the optimal manner[3] according to their incomes, the given price systems, and their common want constitution. The double-expenditure principle may then be interpreted as follows: Group A in price situation 0 is taken to be equally well off as group B in price situation 1 if each would have to cut down (or increase) its total expenditure *in the same proportion* if it should compose its quantity budget in the way the other does. In other words, A says to B: If I were to buy *your* quantity budget q_1 (at my prices p_0), I could cut down (or increase) my total expenditure by x per cent. B replies: If I were to buy *your* quantity budget q_0 (at my prices p_1), I could cut down (or increase) my total expenditure by y per cent. If now $x=y$ (taking account of the signs), Frisch says that the groups A and B live approximately on the same level of well-being. Indeed

$$\frac{\sum p_1q_1 - \sum p_1q_0}{\sum p_1q_1} = \frac{\sum p_0q_0 - \sum p_0q_1}{\sum p_0q_0}$$

may also be written

$$\frac{\sum p_1q_0}{\sum p_1q_1} = \frac{\sum p_0q_1}{\sum p_0q_0}$$

[1] Ragnar Frisch, "Annual Survey of General Economic Theory: The Problem of Index Numbers," ECONOMETRICA, Vol. 4, Jan. 1936, particularly pp. 29–30; *Methods of Measuring the Relative Cost of Living;* Washington, D. C., July 27, 1937 (mimeographed); "The Double-Expenditure Method," ECONOMETRICA, Vol. 6, Jan., 1938, pp. 85–90.

[2] Quantity budget means the combination of quantities of different consumption goods or services which is obtained for the total expenditure.

[3] Optimal composition of the quantity budget means the highest satisfaction with the given total money expenditure.

or $\sum p_0q_0 \sum p_1q_0 = \sum p_1q_1 \sum p_0q_1$. Here the left-hand side is Frisch's double expenditure in the 0-situation (D_0), and the right-hand side the double expenditure in the 1-situation (D_1).*

* Another interpretation is simply to say that the two quantity combinations q_0 and q_1 shall be "equally large" judged by the "ideal" quantity index. Indeed, the above equality can also be written

$$\sqrt{\frac{\sum q_1p_0}{\sum q_0p_0} \cdot \frac{\sum q_1p_1}{\sum q_0p_1}} = 1.$$

This criterion is unaltered if the points 0 and 1 are interchanged, and therefore defines a correspondence between points on the two expansion paths. (This would not have been the case if we had used a quantity index of, say, the Laspeyres or Paasche type.) In terms of the *indicator* the criterion can probably not be justified in any simpler way than that explained in the ECONOMETRICA survey.—R.F.

FINAL ANNOUNCEMENT OF THE KRAKÓW MEETING, SEPTEMBER 18–21, 1938

THE EIGHTH European meeting of the Econometric Society will be held at Kraków, Poland, from September 18 to 21, 1938.

Accommodation with full board will be available at a charge of 10, 15, or 20 zlotys per day. Members desiring to attend the meeting are requested to write before September 1st to Dr. M. J. Ziomek, Towarzystwo Ekonomiczne, Gołębia 20, Kraków, stating the class of accommodation they require and the date of their arrival. They will receive in reply further details regarding programme, accommodation, etc.

The first session will meet on Sunday morning (September 18), but it is proposed that members should meet for the first time on Saturday evening (September 17).

The programme of the meeting will concentrate on three topics: (1) The theory of interest, (2) The theory of production index numbers, (3) The theory of economic planning. Drafts or summaries of papers should be sent before August 1 to Professor W. Zawadzki, Odyńca 9, Warszawa. The detailed programme of the meeting (including the list of papers) will be sent to members desiring to attend on request to Dr. M. J. Ziomek.

W. ZAWADZKI

MEETING OF FELLOWS

The annual meeting of the Fellows of the Econometric Society required by the Constitution will be held at Colorado Springs, Colorado, U. S. A., on Monday, July 18, 1938.

ALFRED COWLES 3RD
Secretary

SUGGESTIONS FOR FELLOWSHIP

The constitution of the Econometric Society states:

All Fellows of the Society shall be nominated by the Council and elected by mail-vote of the Fellows. Such nomination may be made at any time. To be eligible for such nomination a person must have published original contributions to economic theory, or to such statistical, mathematical, or accounting analyses as have a definite bearing on problems in economic theory, and must have been a member of the Society for at least one year. Each year the Council shall offer the members an opportunity to suggest nominees for fellowships.

In accordance with this provision all members of the Econometric Society are hereby invited to suggest nominees for fellowship. Each

nominee should possess the following qualifications laid down by the Council:

1. He should be an economist.
2. He should be a statistician.
3. He should have some knowledge of higher mathematics.
4. He should have made original contributions to economic theory.

The present Fellows are listed in ECONOMETRICA, Vol. 6, January, 1938, p. 95. The membership list published in ECONOMETRICA, Vol. 5, October, 1937, pp. 393–409, may be found convenient in determining those who are eligible for nomination in the present election.

Suggestions for fellowship, accompanied by biographical data and bibliographies of candidates, should be sent to Alfred Cowles 3rd, Secretary of the Econometric Society, 301 Mining Exchange Building, Colorado Springs, Colorado, U.S.A.

MEMBERSHIP LIST CHANGES

A new directory of members will be published in the October issue of ECONOMETRICA. Any errors in the October, 1937 list, or changes in address, should be reported at once to Alfred Cowles 3rd, Secretary of the Econometric Society, 301 Mining Exchange Building, Colorado Springs, Colorado, U. S. A.